



Information Theory

Lecture Notes

Stefan M. Moser

© Copyright Stefan M. Moser

Signal and Information Processing Lab
ETH Zurich
Zurich, Switzerland

Institute of Communications Engineering
National Yang Ming Chiao Tung University (NYCU)
Hsinchu, Taiwan

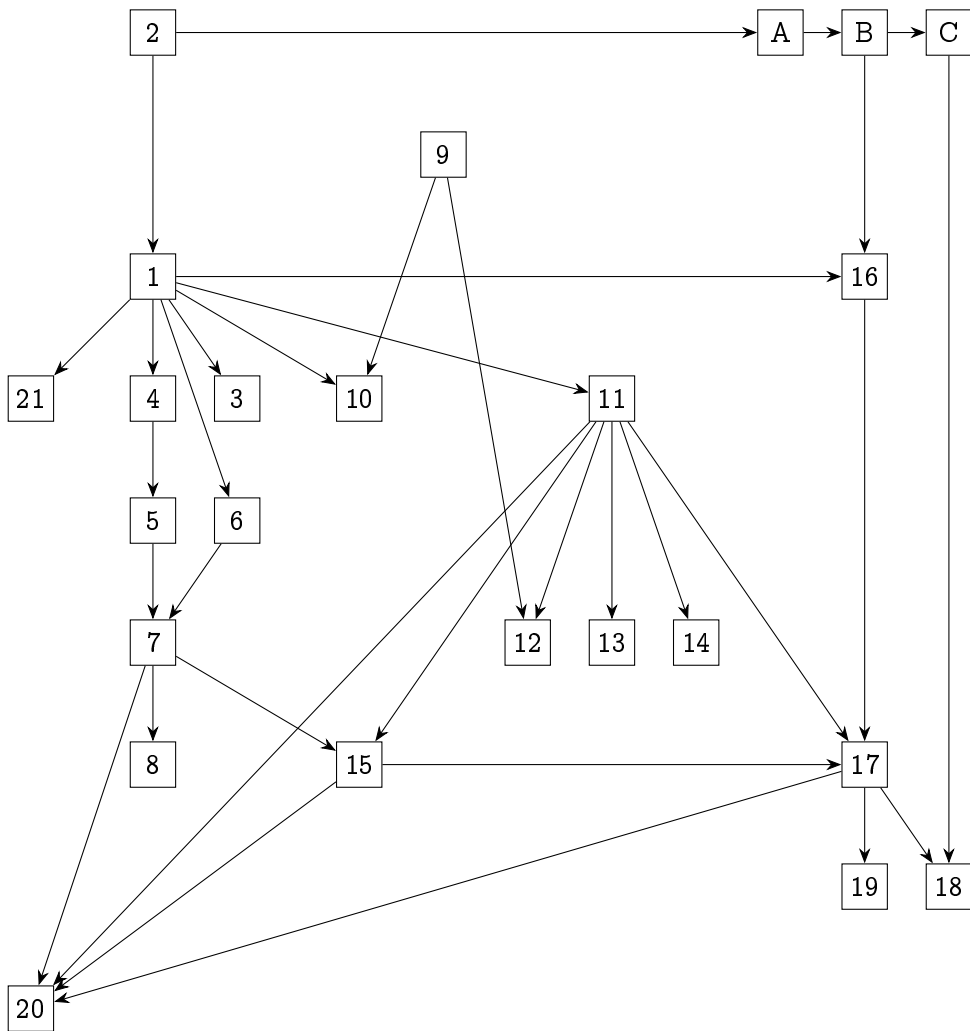
You are welcome to use these lecture notes for yourself, for teaching, or for any other noncommercial purpose. If you use extracts from these lecture notes, please make sure to show their origin. The author assumes no liability or responsibility for any errors or omissions.

6th Edition — 2018.

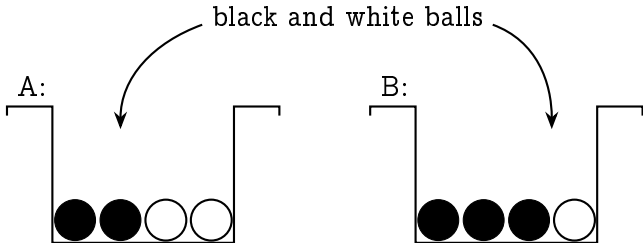
Version 6.21.

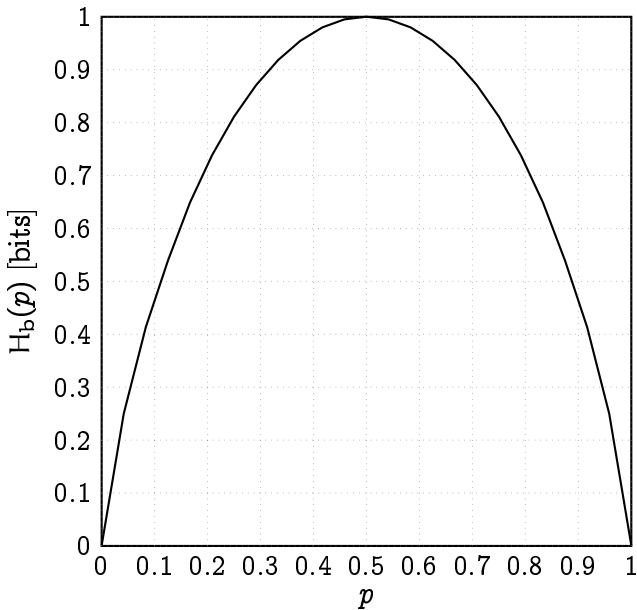
Compiled on 18 December 2025.

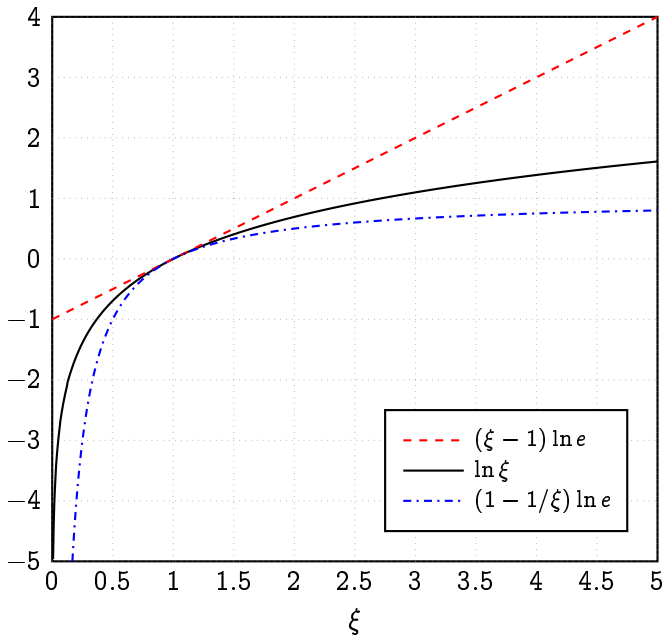
For the latest version see <https://moser-isi.ethz.ch/scripts.html>

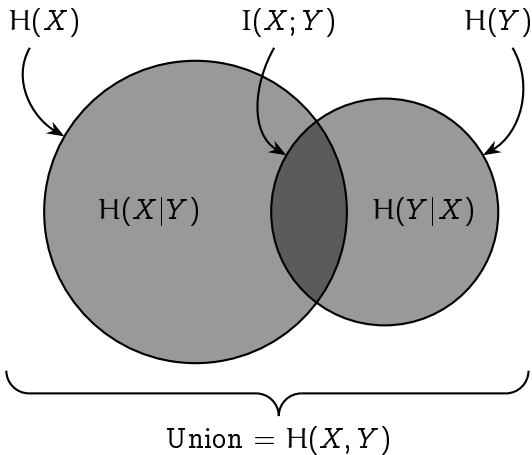


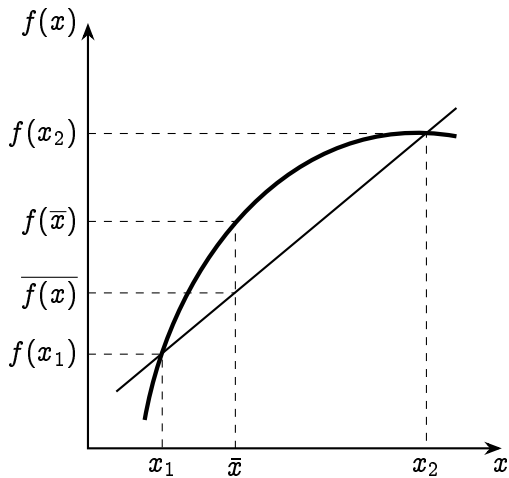
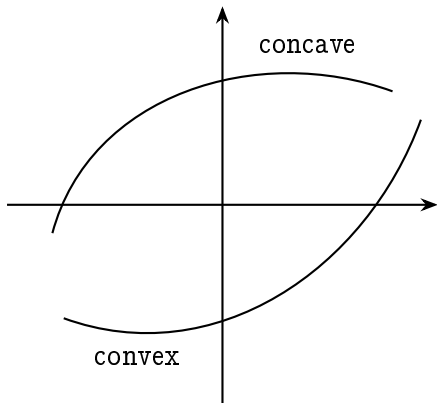
Chapter dependency chart. An arrow has the meaning of “is required for”.

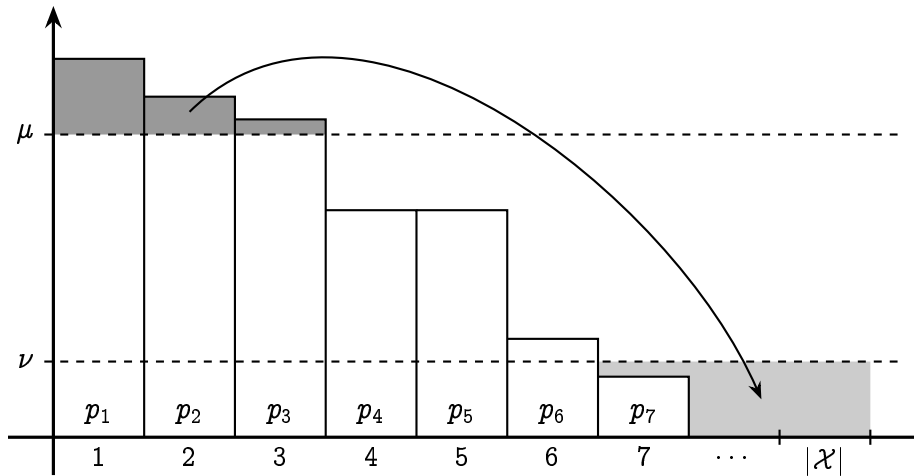


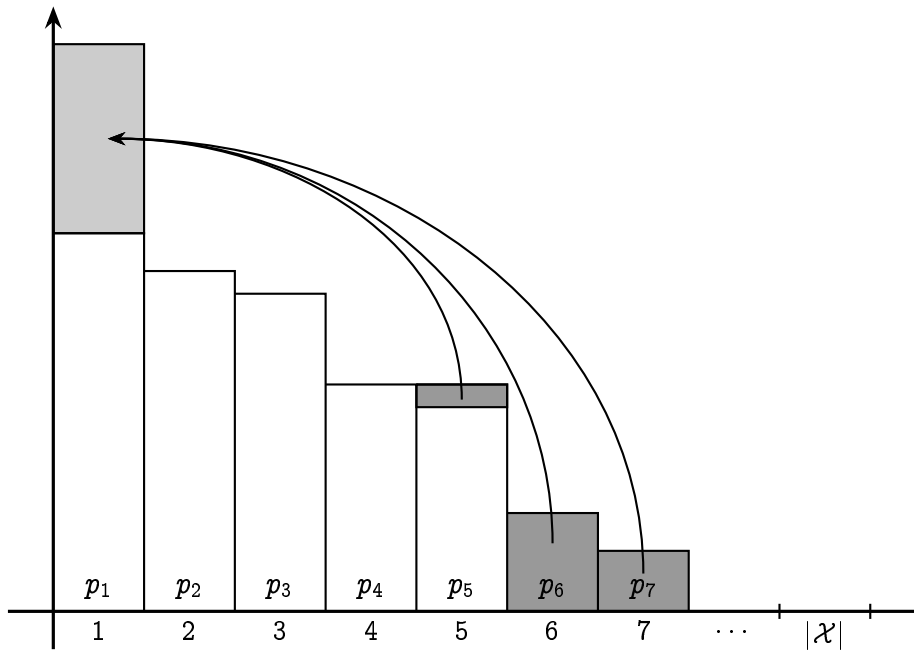




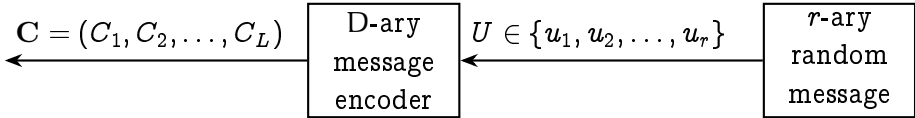




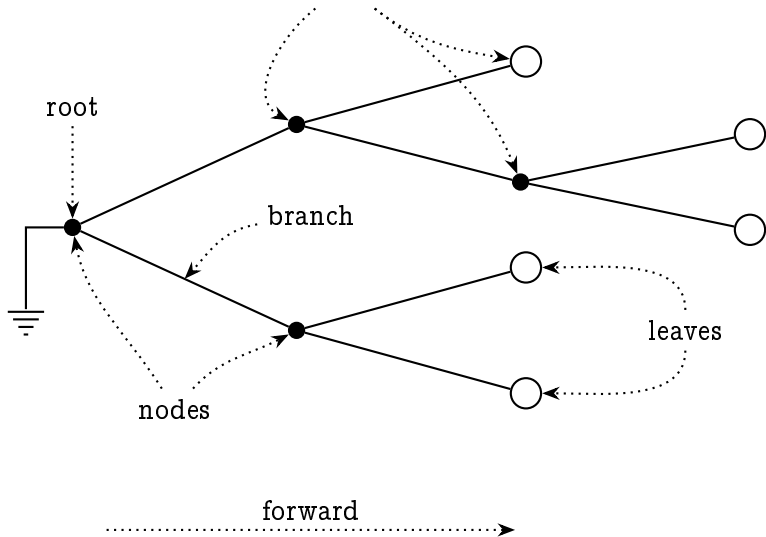


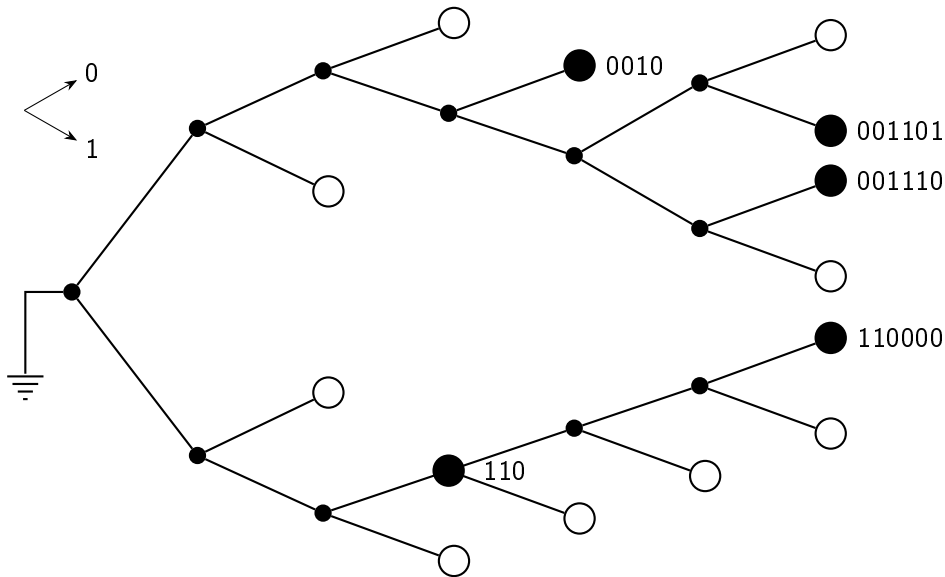


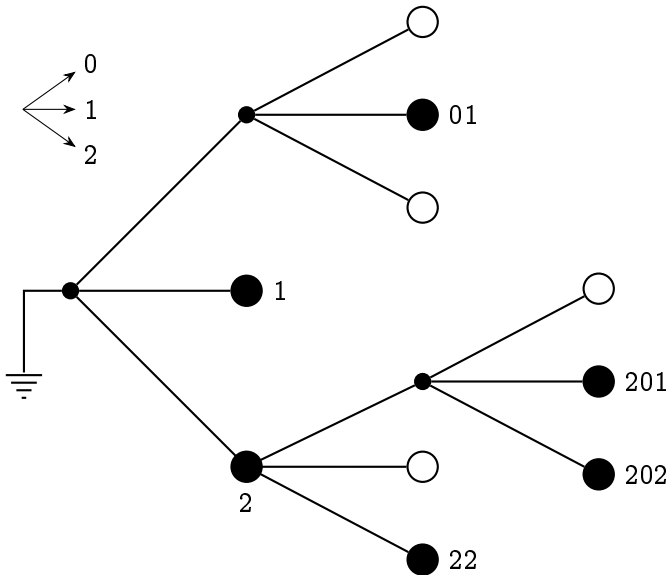
Friend		Alice	Bob	Carol
Probability		$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Phone numbers	(i)	0011	0011	1100
	(ii)	001101	001110	110000
	(iii)	0	1	10
	(iv)	00	11	10
	(v)	0	11	10
	(vi)	10	0	11

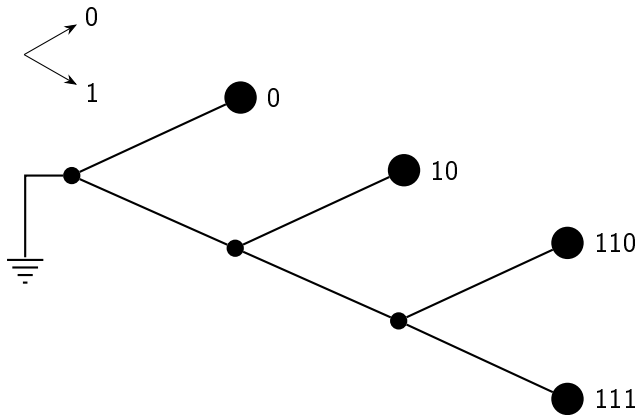


parent node with two children

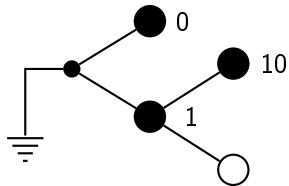






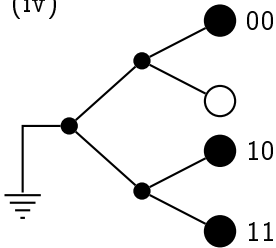


(iii)



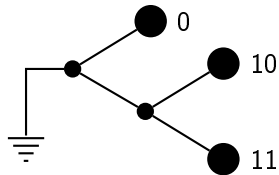
not prefix-free

(iv)

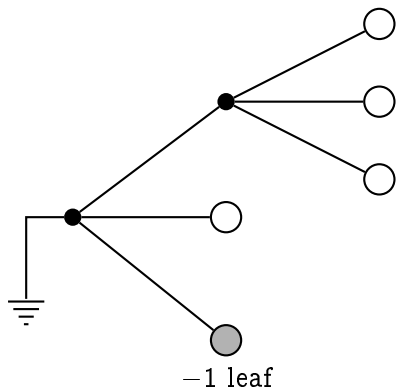


prefix-free

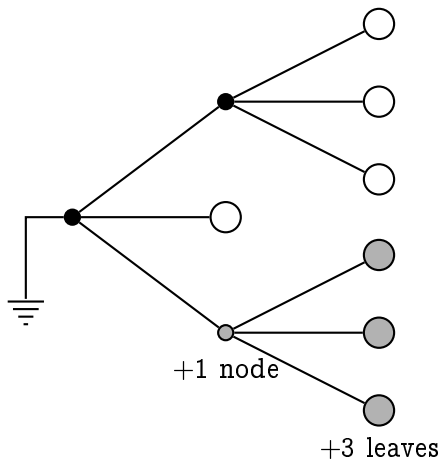
(v)-(vi)

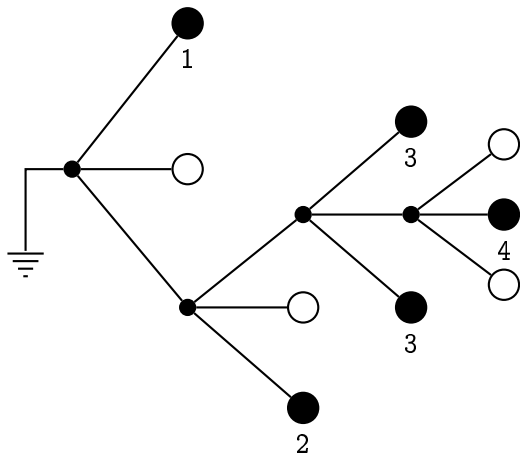
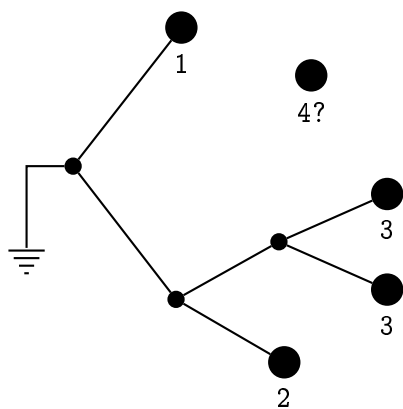
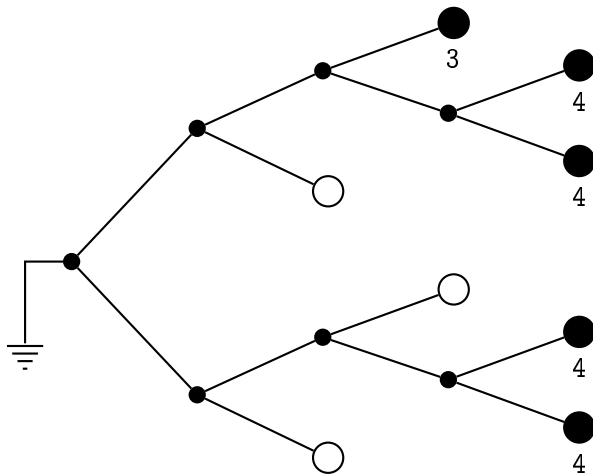


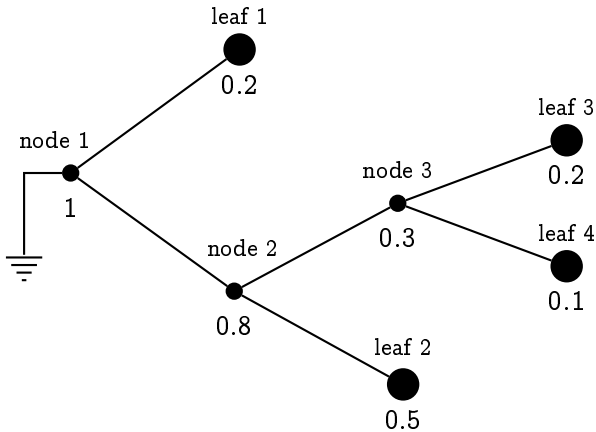
prefix-free

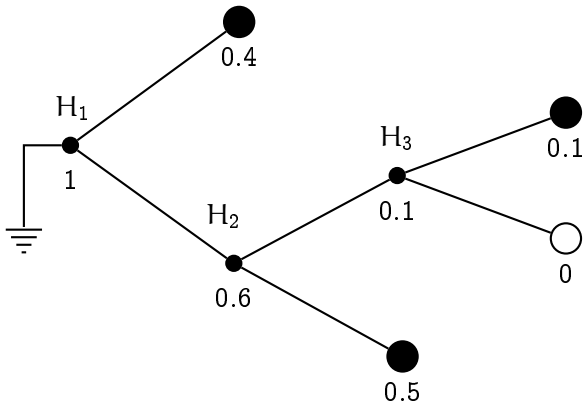


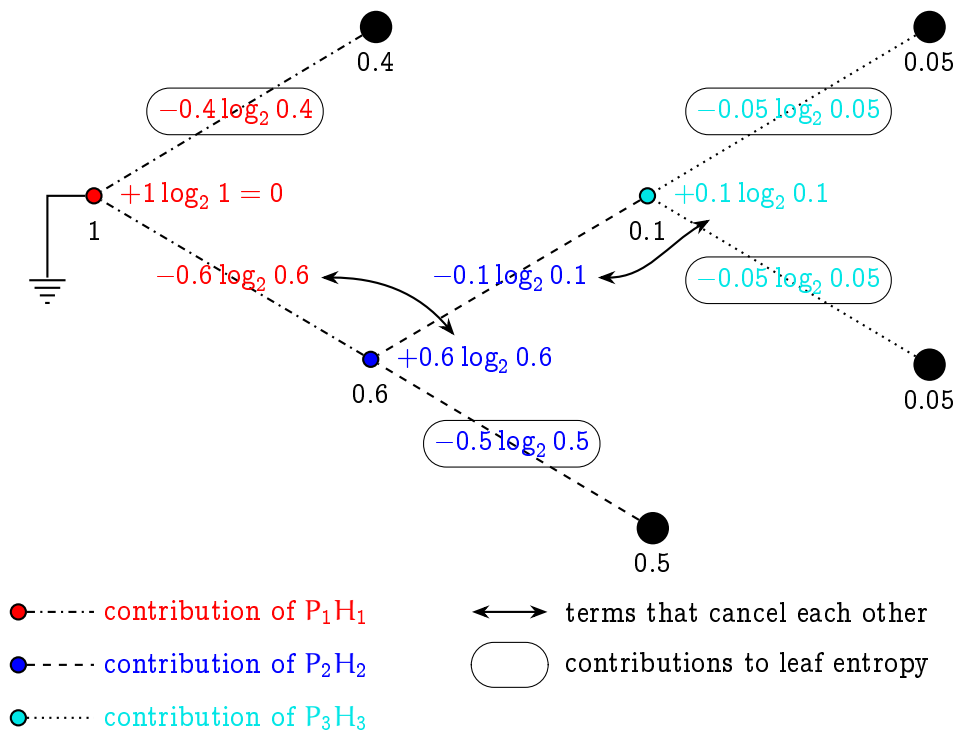
$\Rightarrow$

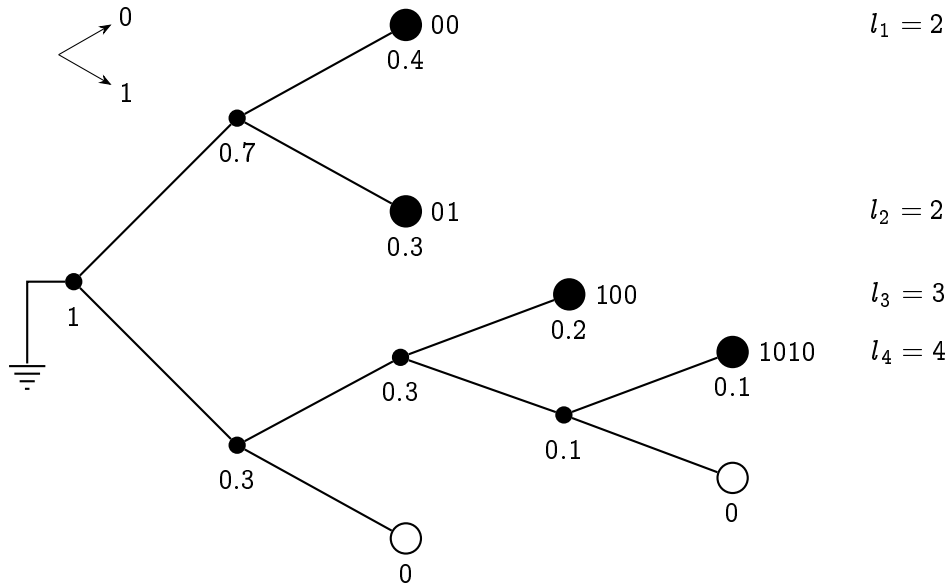


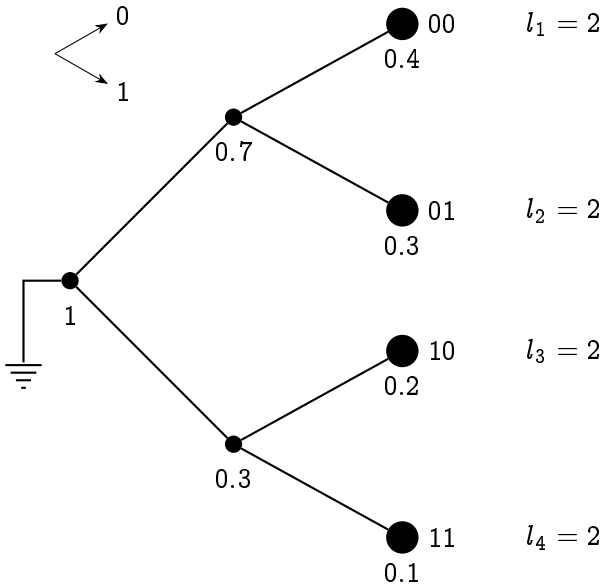












	u_1	u_2	u_3	u_4
p_i	0.4	0.3	0.2	0.1
F_i	0	0.4	0.7	0.9
binary representation	0.0	0.01100...	0.10110...	0.11100...
$\left\lceil \log_2 \frac{1}{p_i} \right\rceil$	2	2	3	4
shortened representation	0.00	0.01	0.101	0.1110
$\mathbf{c}_i$	00	01	101	1110

	u_1	u_2	u_3	u_4
p_i	0.4	0.3	0.2	0.1
F_i	0	0.4	0.7	0.9
ternary representation	0.0	0.10121...	0.20022...	0.22002...
$\left\lceil \log_3 \frac{1}{p_i} \right\rceil$	1	2	2	3
shortened representation	0.0	0.10	0.20	0.220
$\mathbf{c}_i$	0	10	20	220

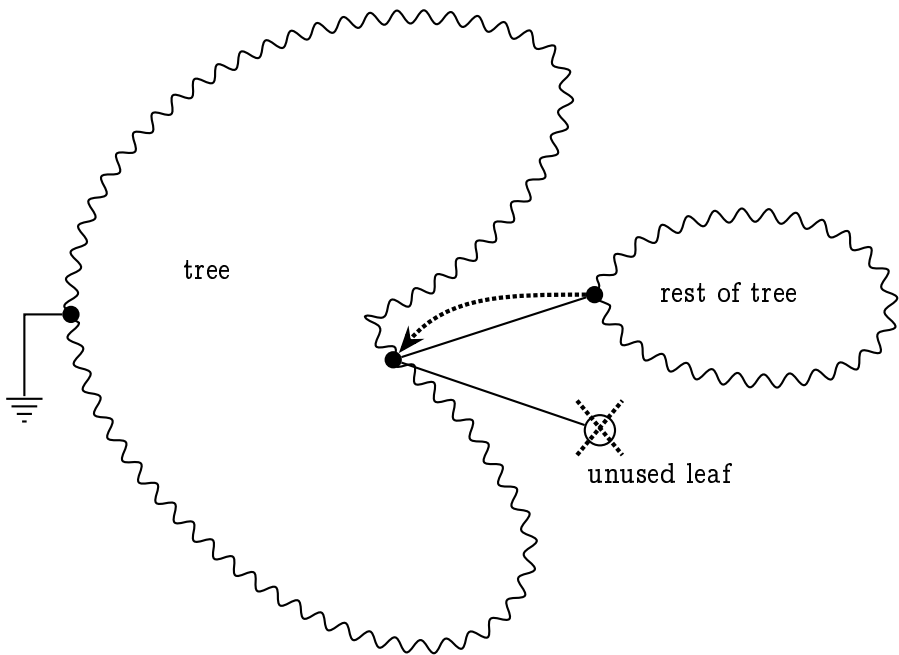
p_1	p_2	p_3	p_4	p_5
0.35	0.25	0.15	0.15	0.1
0.6		0.4		
0		1		
0.35	0.25	0.15	0.15	0.1
0	1	0.15	0.25	
			1	
			0.15	0.1
			0	1
00	01	10	110	111

p_1	p_2	p_3	p_4	p_5
0.35	0.25	0.15	0.15	0.1
0.35	0.25	0.4		
0	1	2		
		0.15	0.15	0.1
		0	1	2
0	1	20	21	22

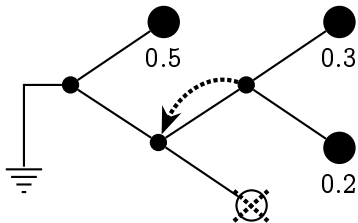
p_1	p_2	p_3	p_4	p_5	p_6	p_7
0.35	0.3	0.15	0.05	0.05	0.05	0.05
0.65 0		0.35 1				
0.35	0.3	0.15	0.05	0.05	0.05	0.05
0	1	0.2 0		0.15 1		
		0.15	0.05	0.05	0.05	0.05
		0	1	0.1 0		0.05 1
				0.05	0.05	
00	01	100	101	1100	1101	111

p_1	p_2	p_3	p_4	p_5	p_6	p_7
0.35	0.3	0.15	0.05	0.05	0.05	0.05
0.35 0	0.65					
	1					
	0.3	0.15	0.05	0.05	0.05	0.05
	0.3 0	0.35				
		1				
		0.15	0.05	0.05	0.05	0.05
		0.15 0	0.2			
			1			
			0.05	0.05	0.05	0.05
			0.1		0.1	
			0		1	
			0.05	0.05	0.05	0.05
			0	1	0	1
0	10	110	11100	11101	11110	11111

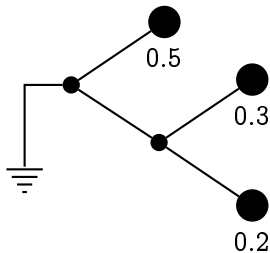
p_1	p_2	p_3	p_4	p_5	p_6	p_7	
0.35	0.3	0.15	0.05	0.05	0.05	0.05	
0.35 0	0.65 1						
	0.3	0.15	0.05	0.05	0.05	0.05	
	0.3 0	0.35 1					
		0.15	0.05	0.05	0.05	0.05	
		0.2 0		0.15 1			
		0.15	0.05	0.05	0.05	0.05	
		0	1	0.1 0		0.05 1	
				0.05	0.05		
				0	1		
	0	10	1100	1101	11100	11101	1111



(i)

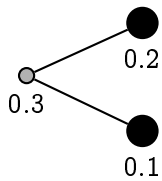


(ii)

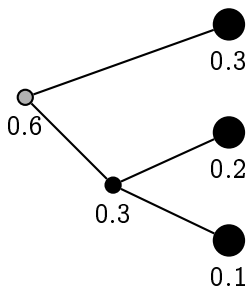




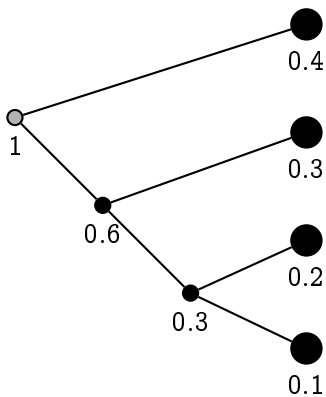
Step 1



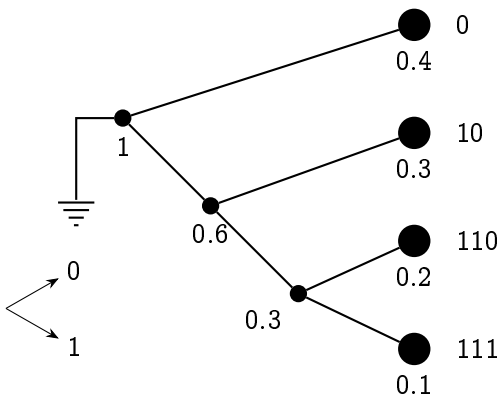
Step 2, first time



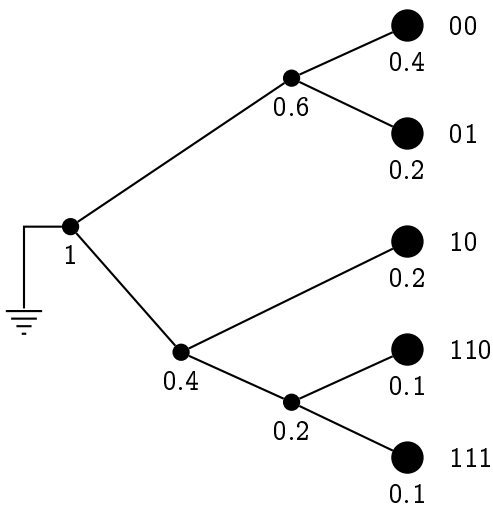
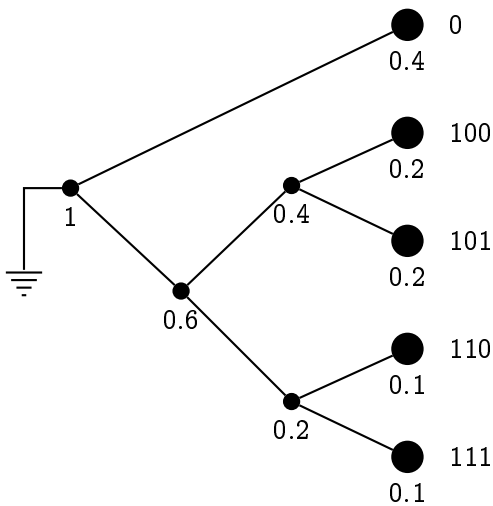
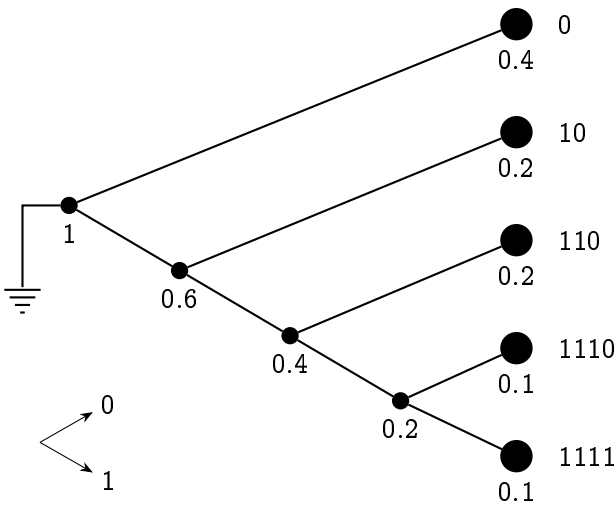
Step 2, second time

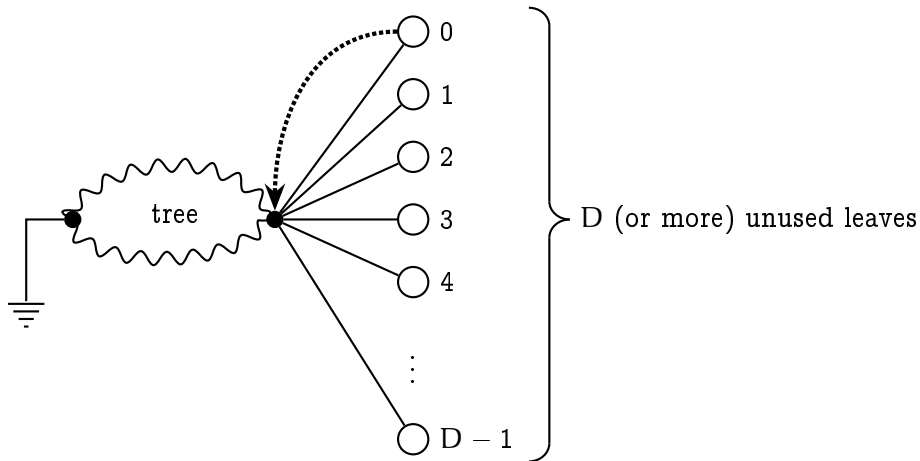
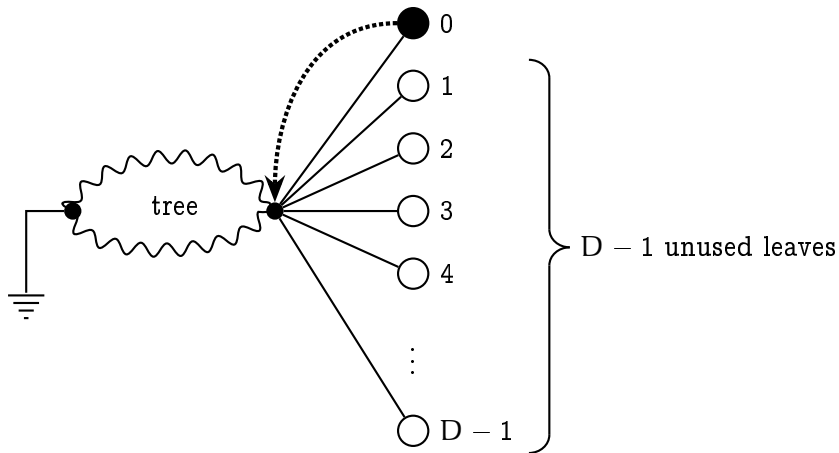


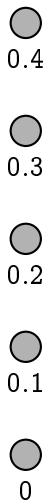
Step 2, third time



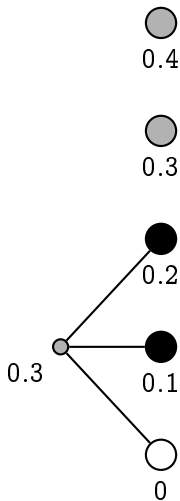
Step 3



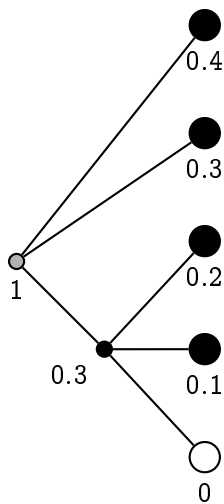




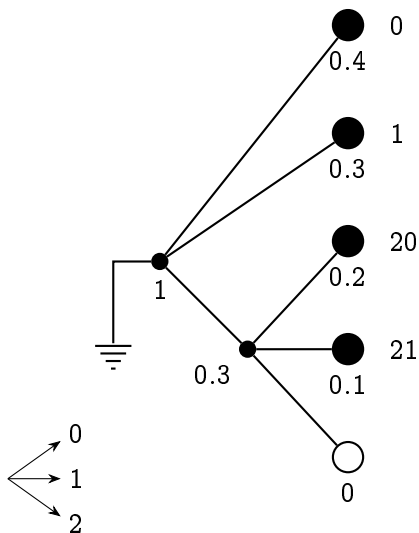
Step 1



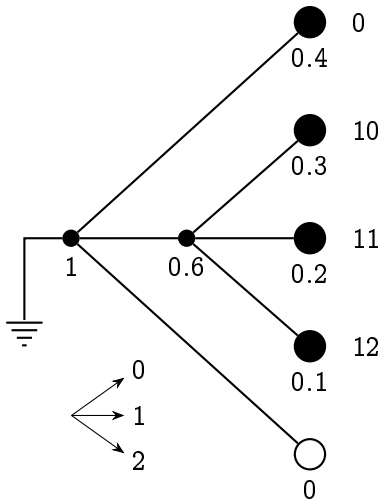
Step 2, first time



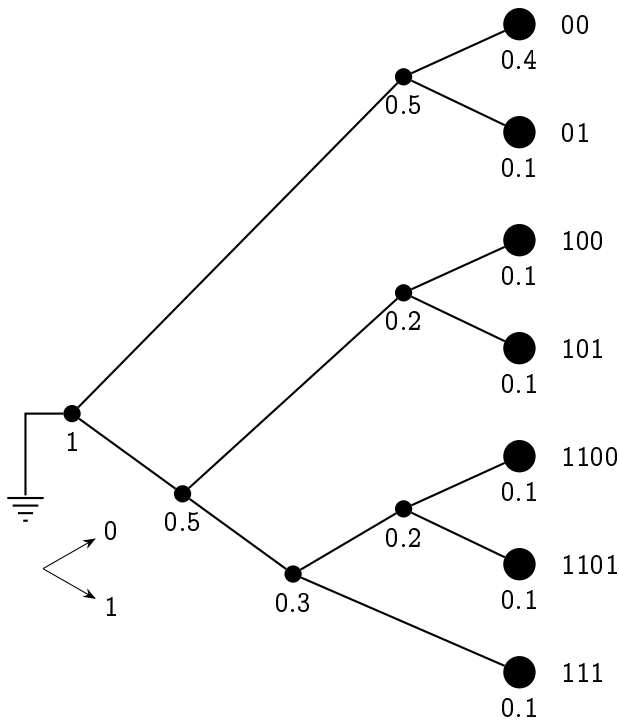
Step 2, second time



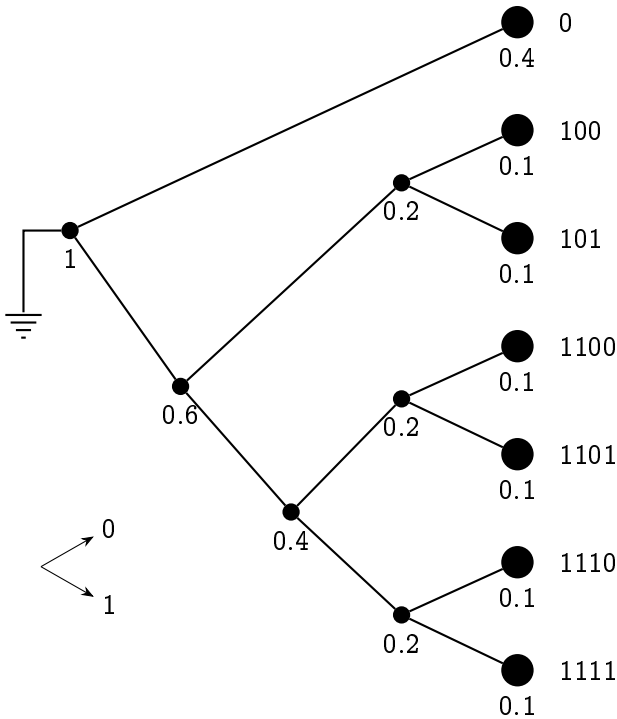
Step 3



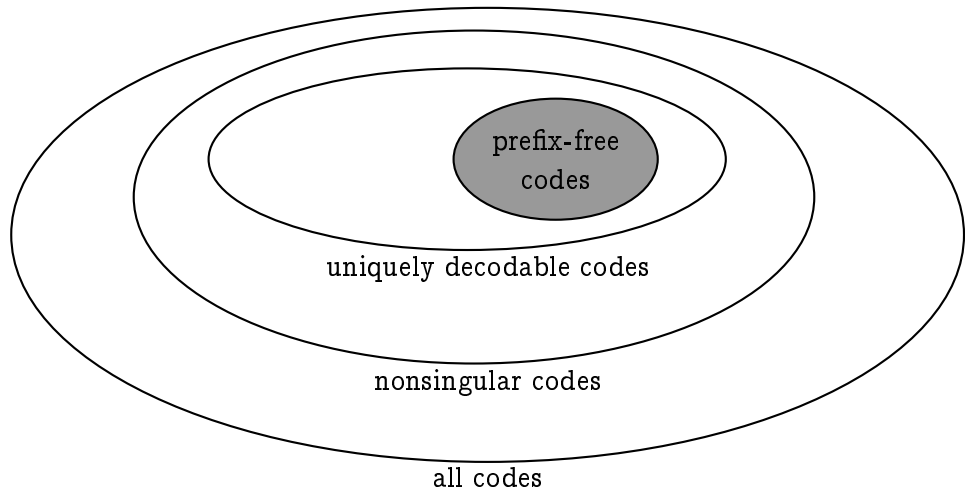
p_1	p_2	p_3	p_4	p_5	p_6	p_7
0.4	0.1	0.1	0.1	0.1	0.1	0.1
0.5 0		0.5 1				
0.4	0.1	0.1	0.1	0.1	0.1	0.1
0	1	0.2 0		0.3 1		
		0.1	0.1	0.1	0.1	0.1
		0	1	0.2 0		0.1 1
				0.1	0.1	
00		01		1100		111
				1101		

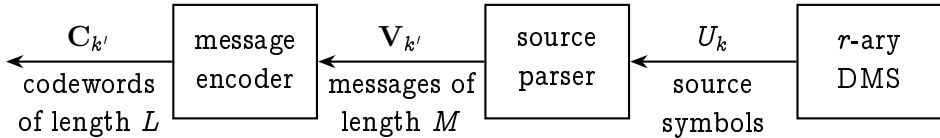


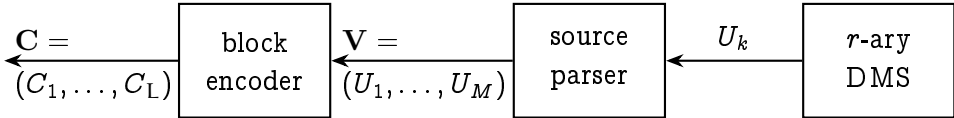
	u_1	u_2	u_3	u_4	u_5	u_6	u_7
p_i	0.4	0.1	0.1	0.1	0.1	0.1	0.1
F_i	0	0.4	0.5	0.6	0.7	0.8	0.9
binary representation	0.0	0.01100	0.1	0.10011	0.10110	0.11001	0.11100
$\left\lceil \log_2 \frac{1}{p_i} \right\rceil$	2	4	4	4	4	4	4
shortened representation	0.00	0.0110	0.1000	0.1001	0.1011	0.1100	0.1110
$\mathbf{c}_i$	00	0110	1000	1001	1011	1100	1110

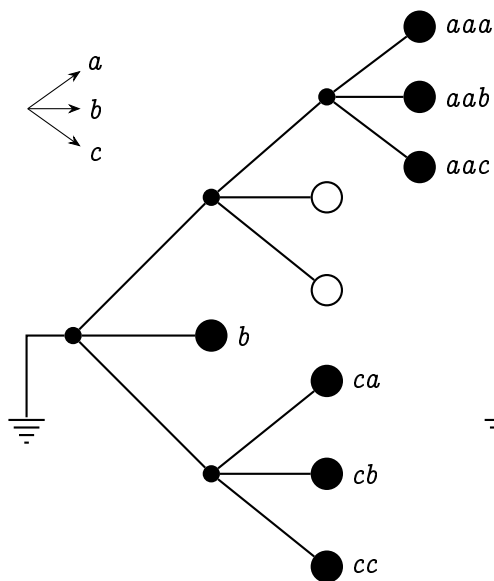


U	Code (i)	Code (ii)	Code (iii)	Code (iv)
a	0	0	10	0
b	0	010	00	10
c	1	01	11	110
d	1	10	110	111

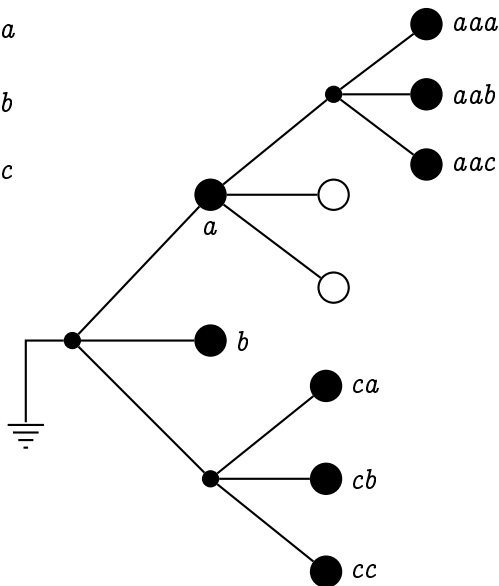




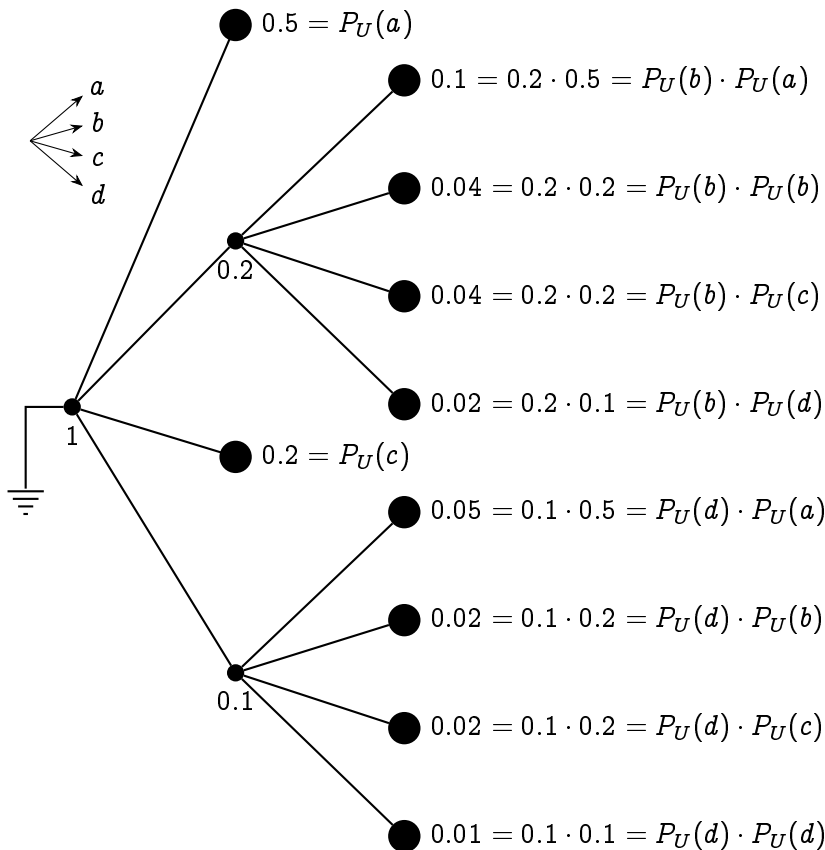


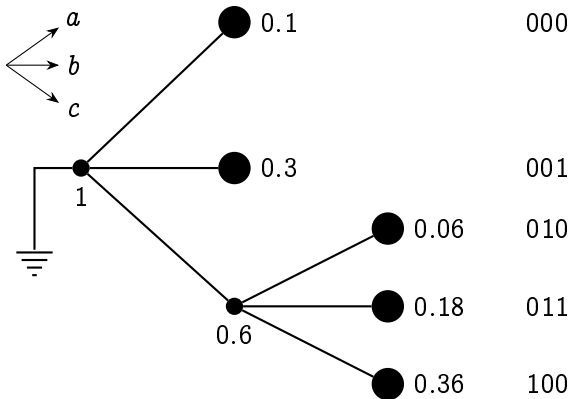


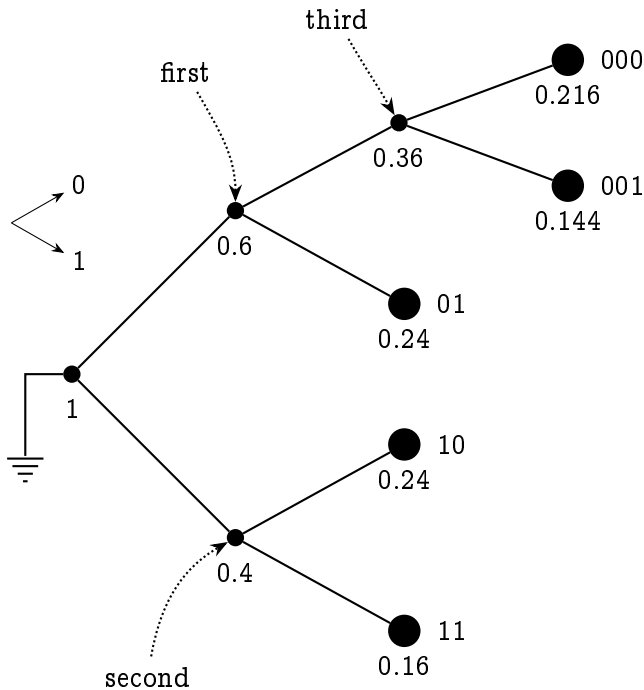
"illegal"

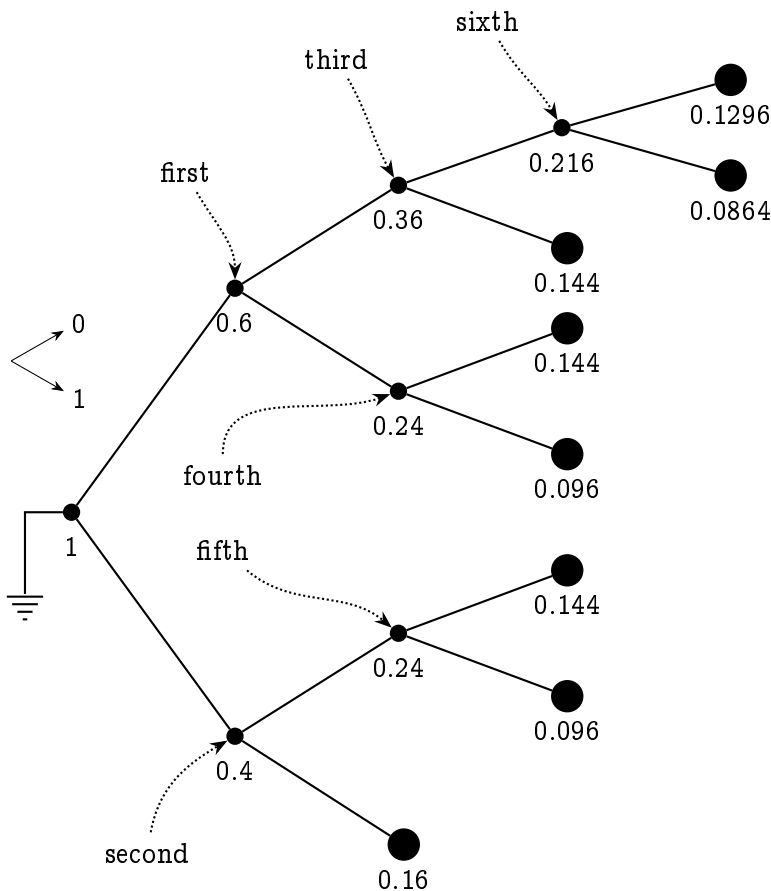


"legal", but not proper

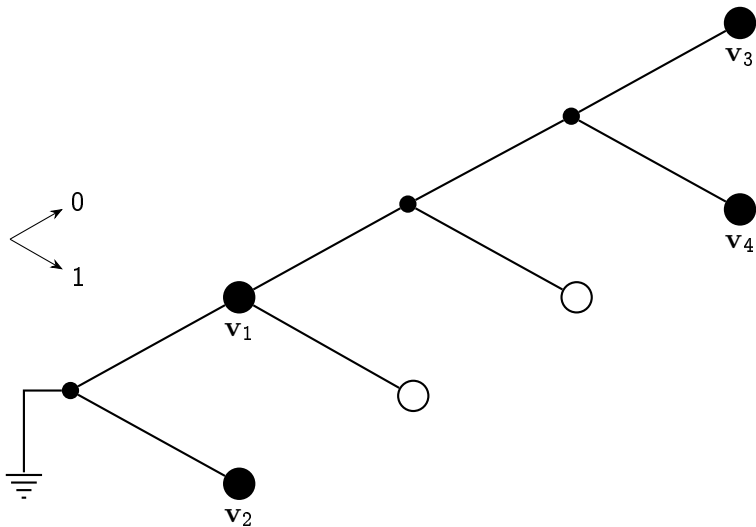


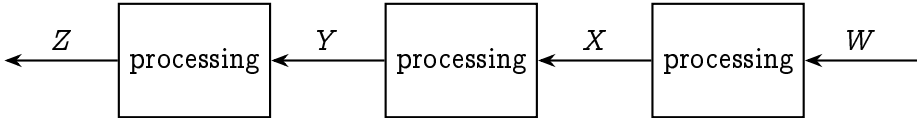


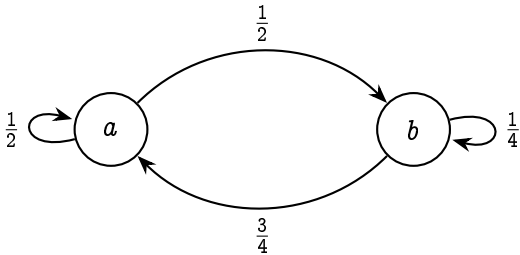


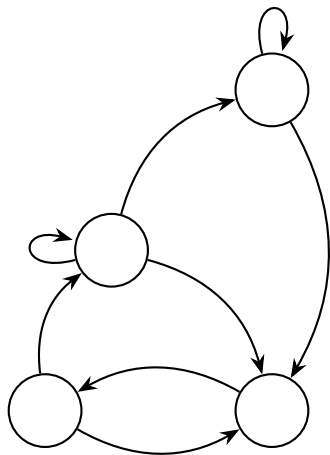


Message	Codewords
0000	000
0001	001
001	010
010	011
011	100
100	101
101	110
11	111

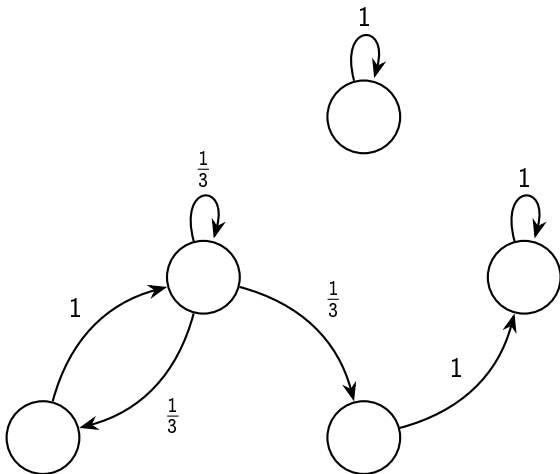




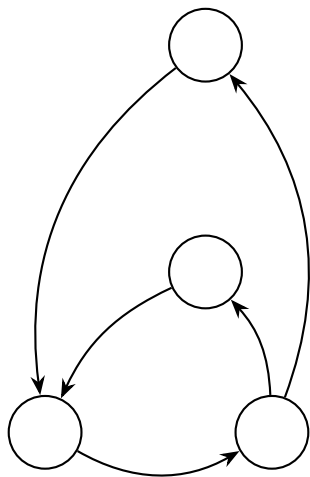




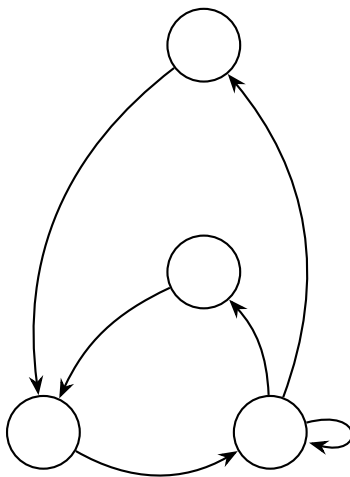
irreducible



reducible



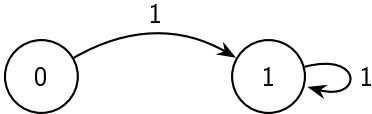
periodic



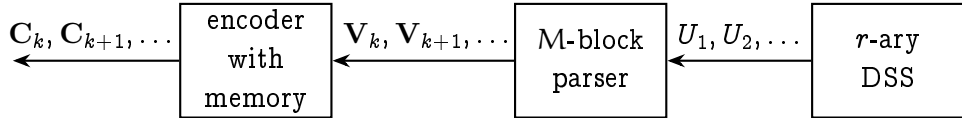
aperiodic

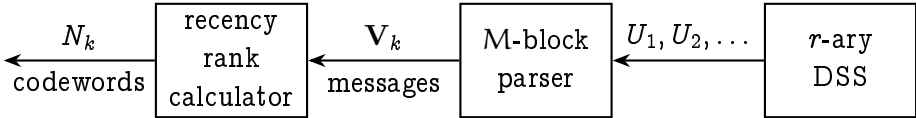
$P_{X_k}(\cdot)$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$\mathbb{P}[X_k = a]$	1	$\frac{1}{2} = 0.5$	$\frac{5}{8} = 0.625$	$\frac{19}{32} = 0.59375$
$\mathbb{P}[X_k = b]$	0	$\frac{1}{2} = 0.5$	$\frac{3}{8} = 0.375$	$\frac{13}{32} = 0.40625$

$P_{X_k}(\cdot)$	$k = 5$	$\dots$	$k = \infty$
$\mathbb{P}[X_k = a]$	$\frac{77}{128} = 0.6015625$	$\dots$	$\frac{3}{5} = 0.6$
$\mathbb{P}[X_k = b]$	$\frac{51}{128} = 0.3984375$	$\dots$	$\frac{2}{5} = 0.4$



Stefan M. Moser, *Information Theory*, Version 6.21.





Message	Recency Rank at Time k
---------	--------------------------

00	4
----	---

01	1
----	---

10	3
----	---

11	2
----	---

Message	Recency Rank at Time $k + 1$
---------	------------------------------

00	4
----	---

01	2
----	---

10	1
----	---

11	3
----	---

Message	Recency Rank at Time 1 (by definition)
---------	---

00	1
----	---

01	2
----	---

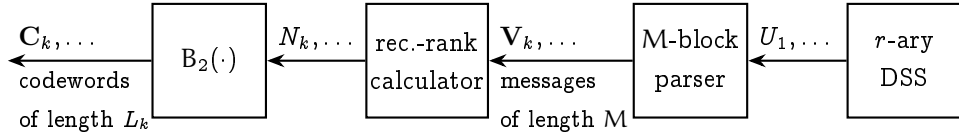
10	3
----	---

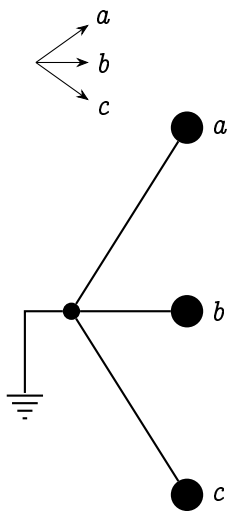
11	4
----	---

n	1	2	3	4	5	6	7	8	9	10	$\dots$
$B_0(n)$	1	10	11	100	101	110	111	1000	1001	1010	$\dots$

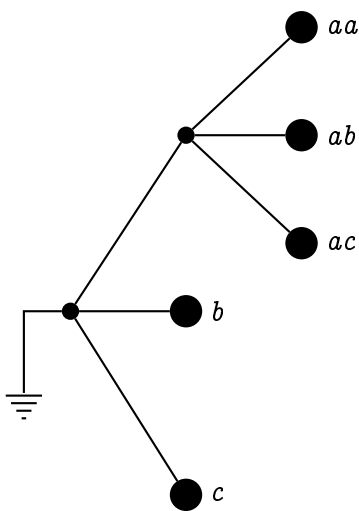
n	1	2	3	4	5	6
$B_0(n)$	1	10	11	100	101	110
$L_{B_0}(n) - 1$	0	1	1	2	2	2
$B_1(n)$	1	010	011	00100	00101	00110
n	7	8	9	10	11	...
$B_0(n)$	111	1000	1001	1010	1011	...
$L_{B_0}(n) - 1$	2	3	3	3	3	...
$B_1(n)$	00111	0001000	0001001	0001010	0001011	...

n	1	2	3	4	5	6
$B_0(n)$	1	10	11	100	101	110
$L_{B_0}(n)$	1	2	2	3	3	3
$B_1(L_{B_0}(n))$	1	010	010	011	011	011
$B_2(n)$	1	0100	0101	01100	01101	01110
n	7	8	9	10	11	...
$B_0(n)$	111	1000	1001	1010	1011	...
$L_{B_0}(n)$	3	4	4	4	4	...
$B_1(L_{B_0}(n))$	011	00100	00100	00100	00100	...
$B_2(n)$	01111	00100000	00100001	00100010	00100011	...

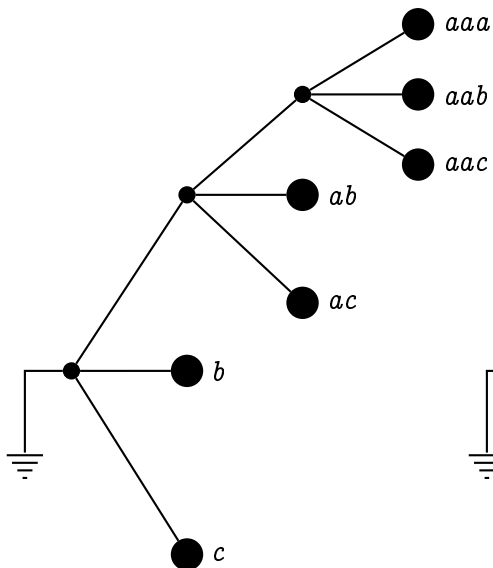




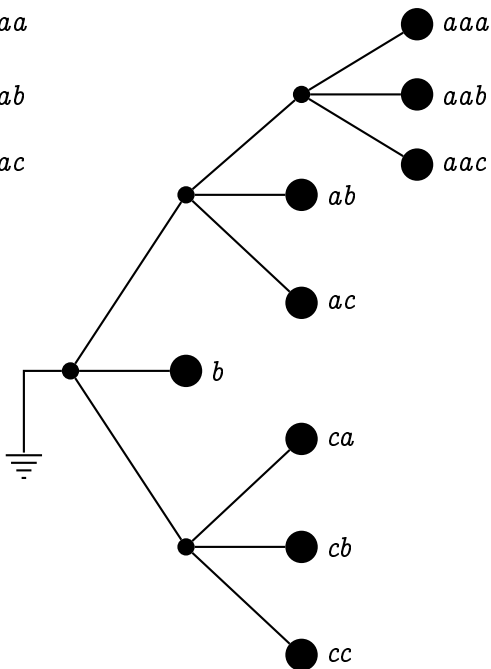
(a)



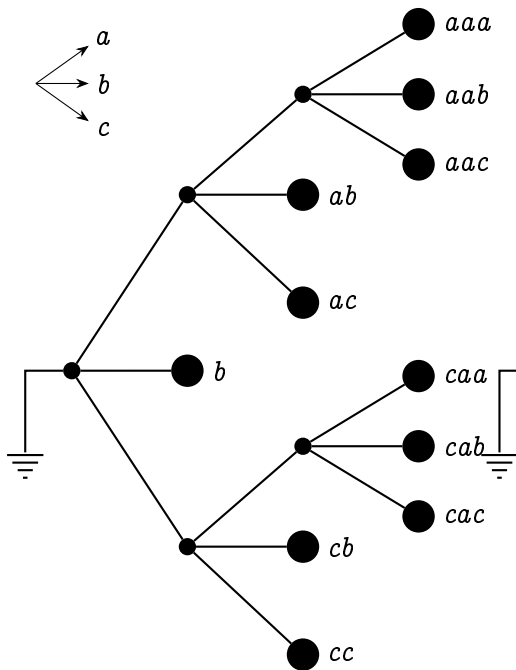
(b)



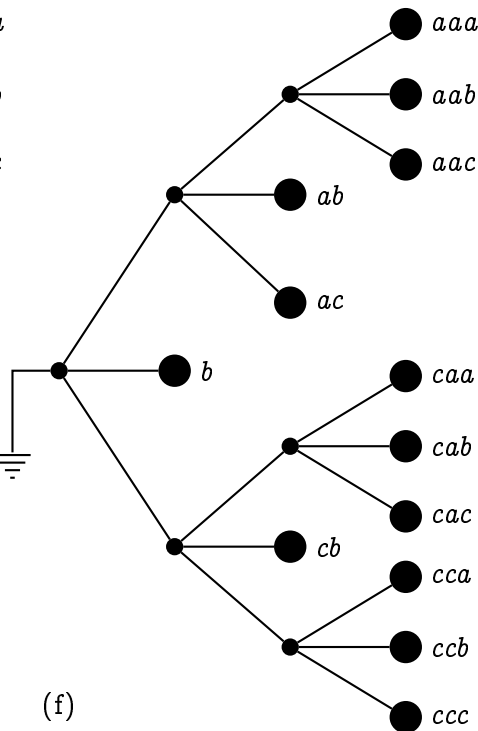
(c)



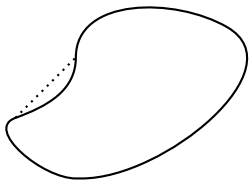
(d)



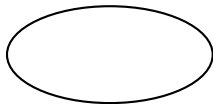
(e)



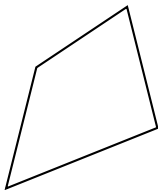
(f)



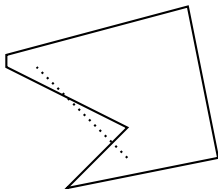
nonconvex



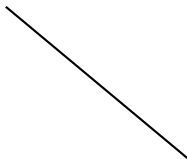
convex



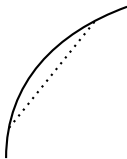
convex



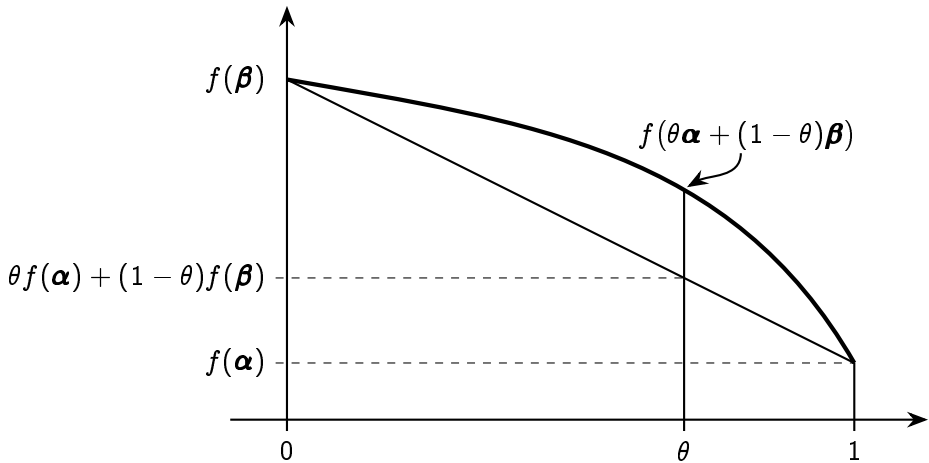
nonconvex

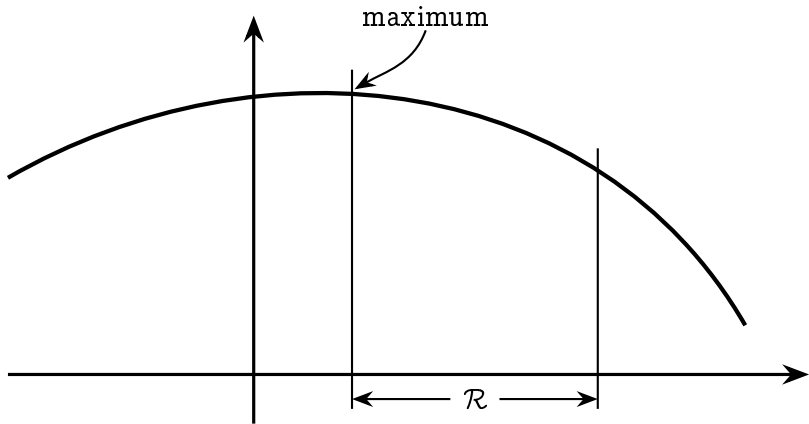


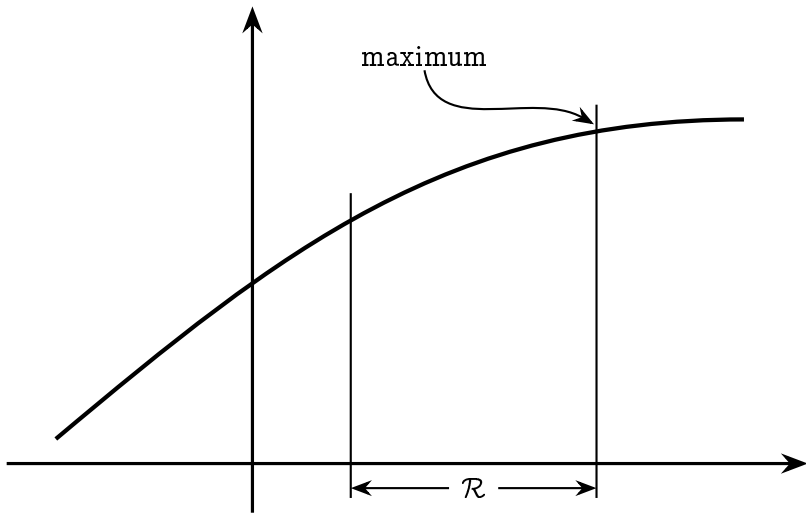
convex



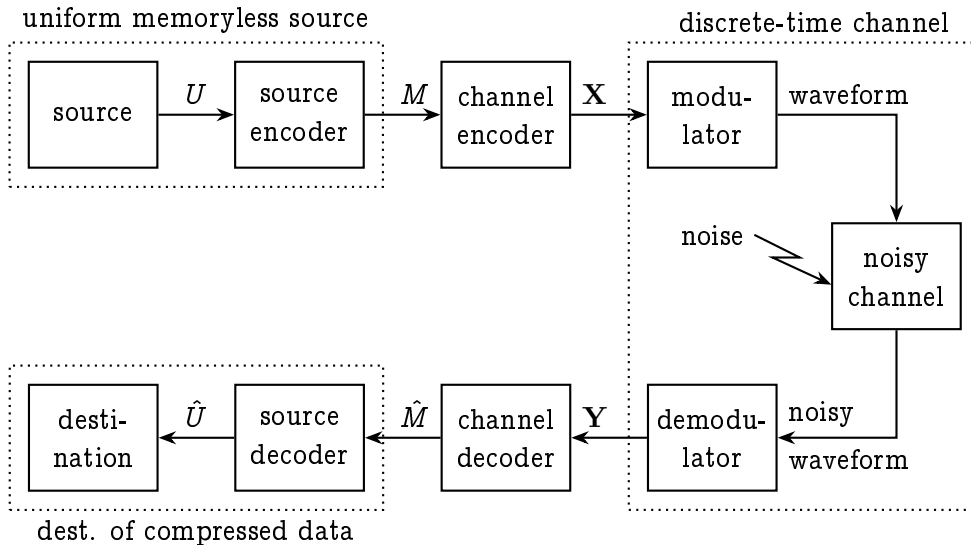
nonconvex

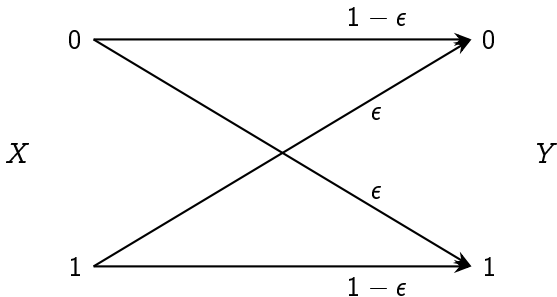


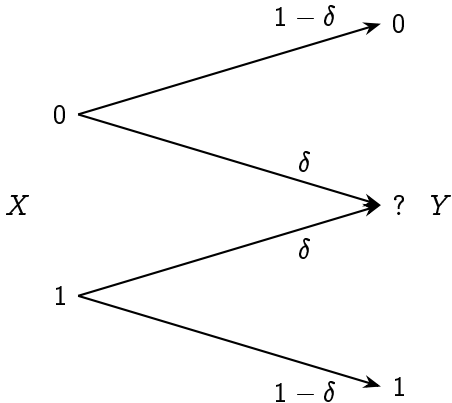


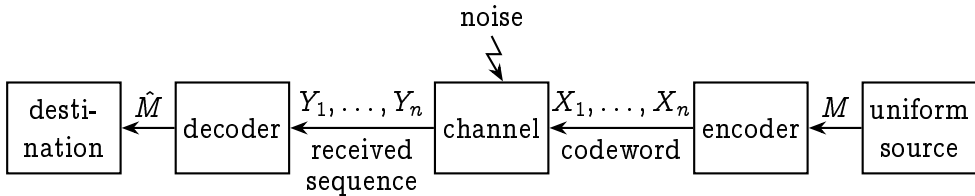


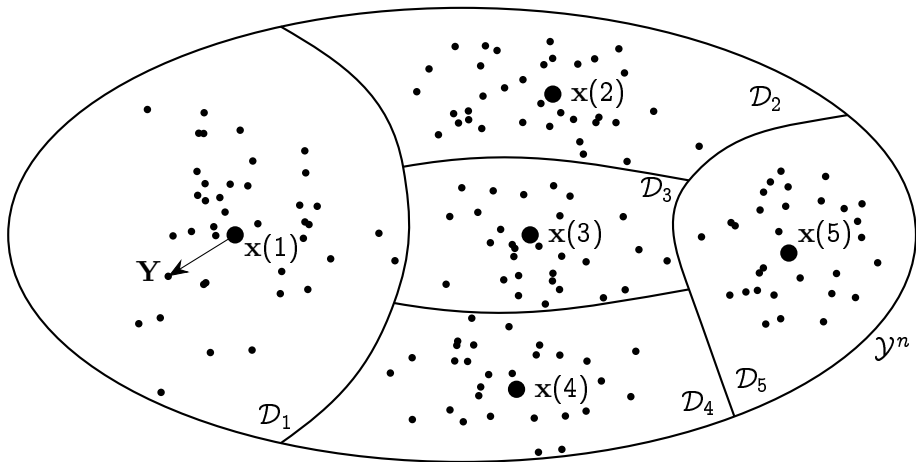
original name	2	3	1
new name t	1	2	3
p_i	0.2	0.1	0.7
o_i	6	10	1.2
$p_i o_i$	1.2	1.0	0.84
	$\vee$	$\vee$	$\wedge$
β_t	1	0.96	0.955

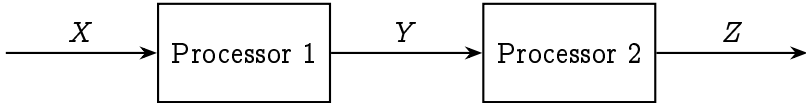




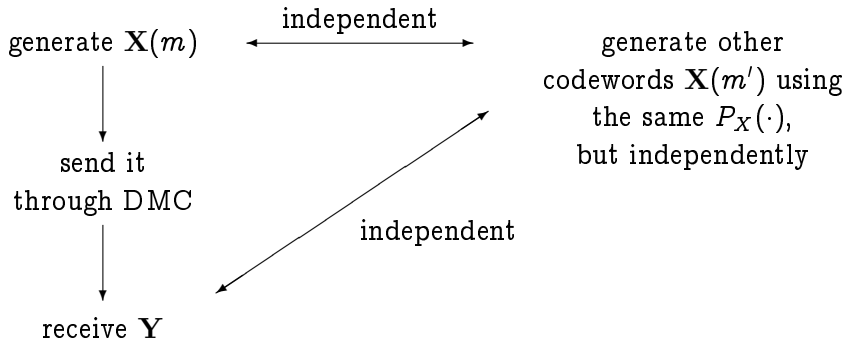


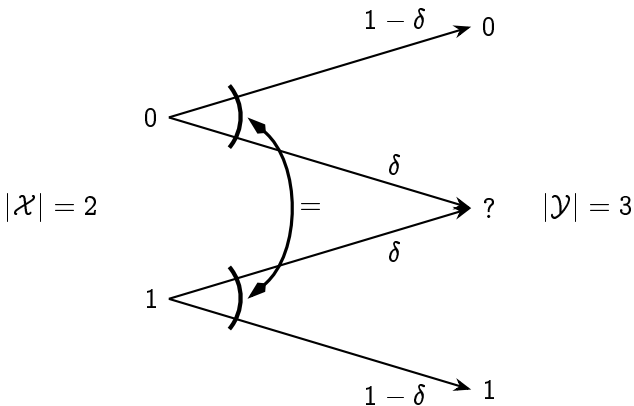


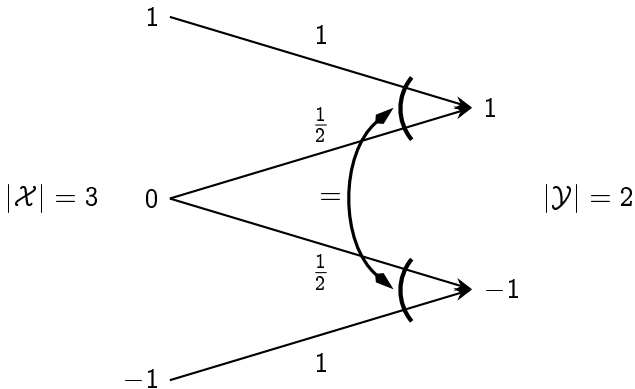


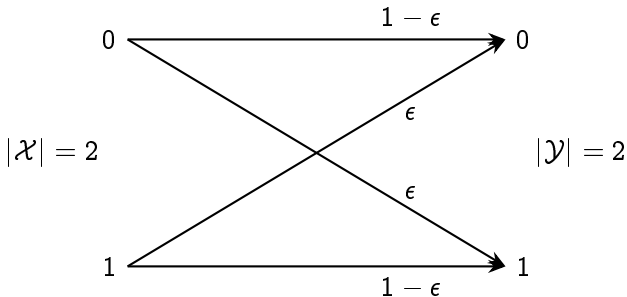


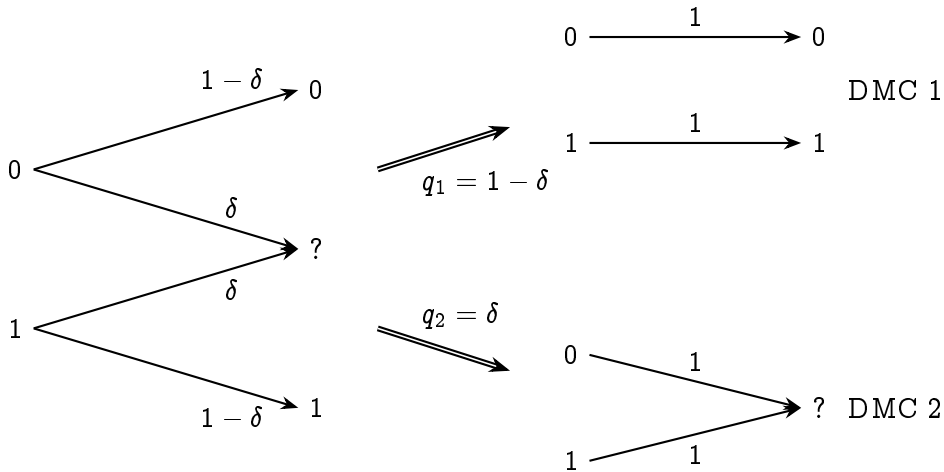
Stefan M. Moser, *Information Theory*, Version 6.21.

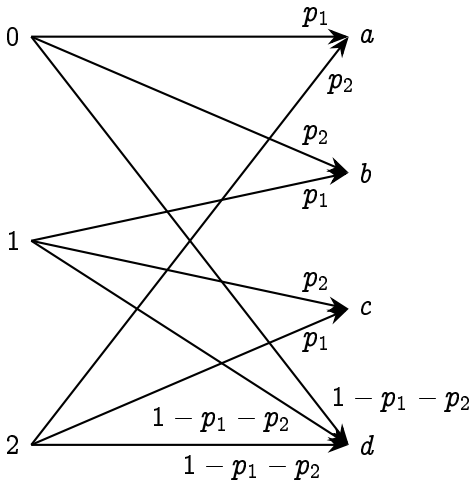


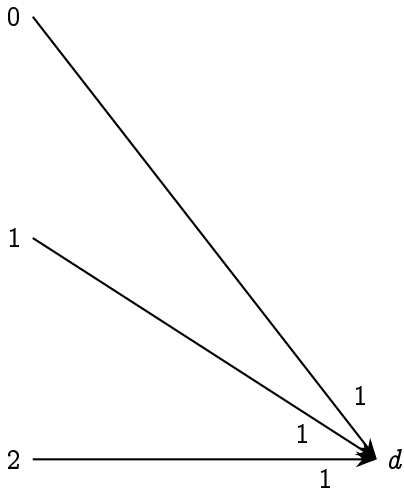
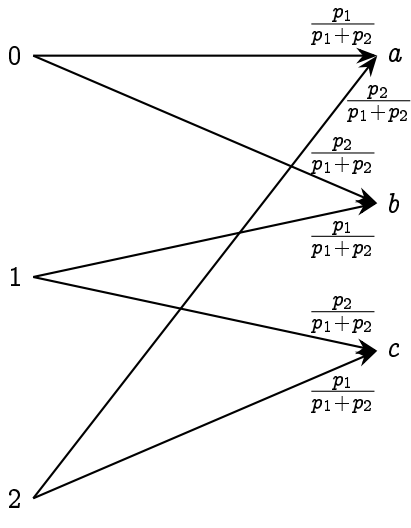


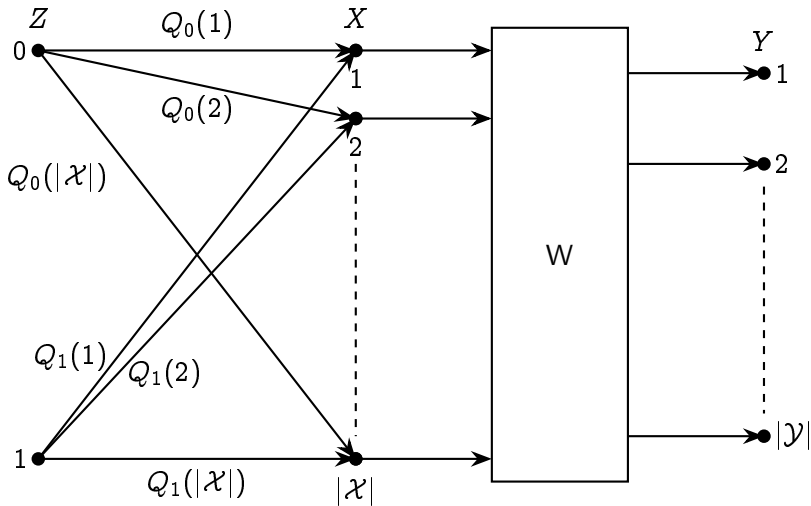




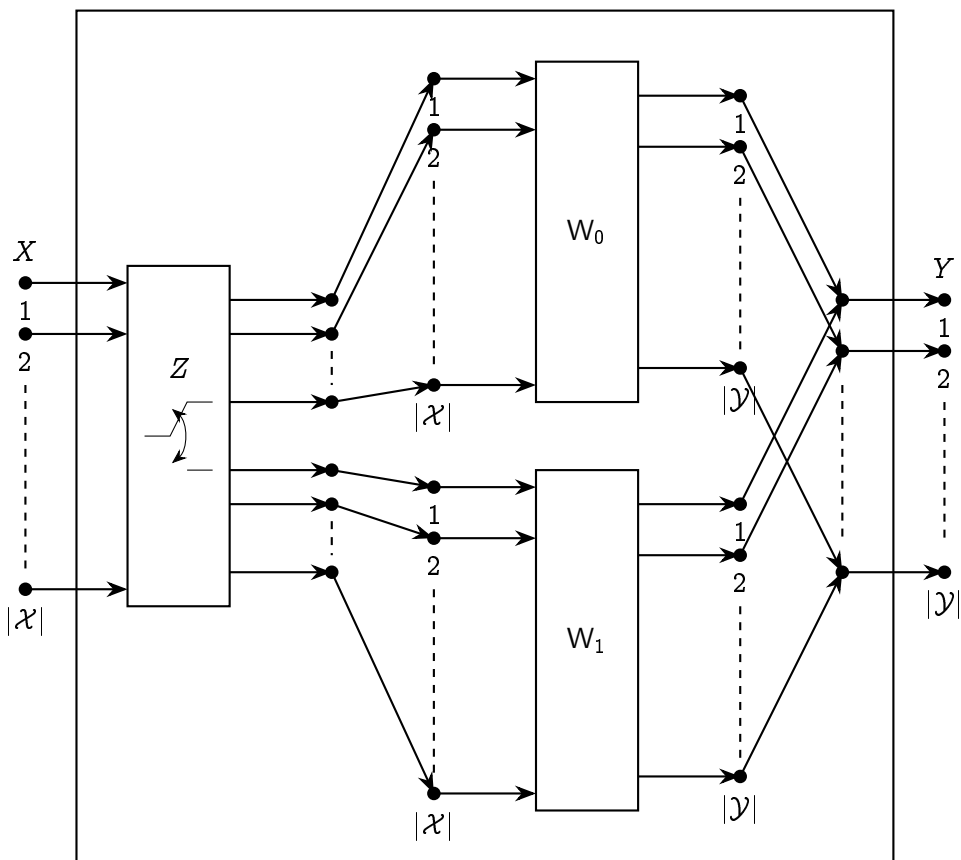


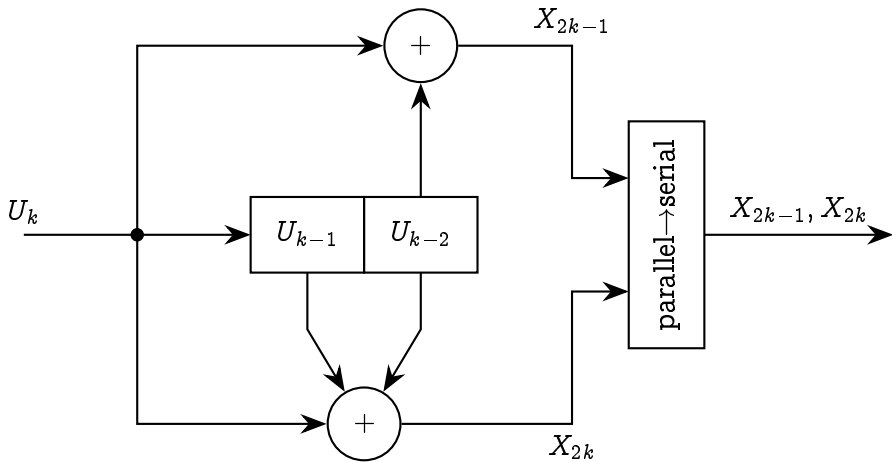


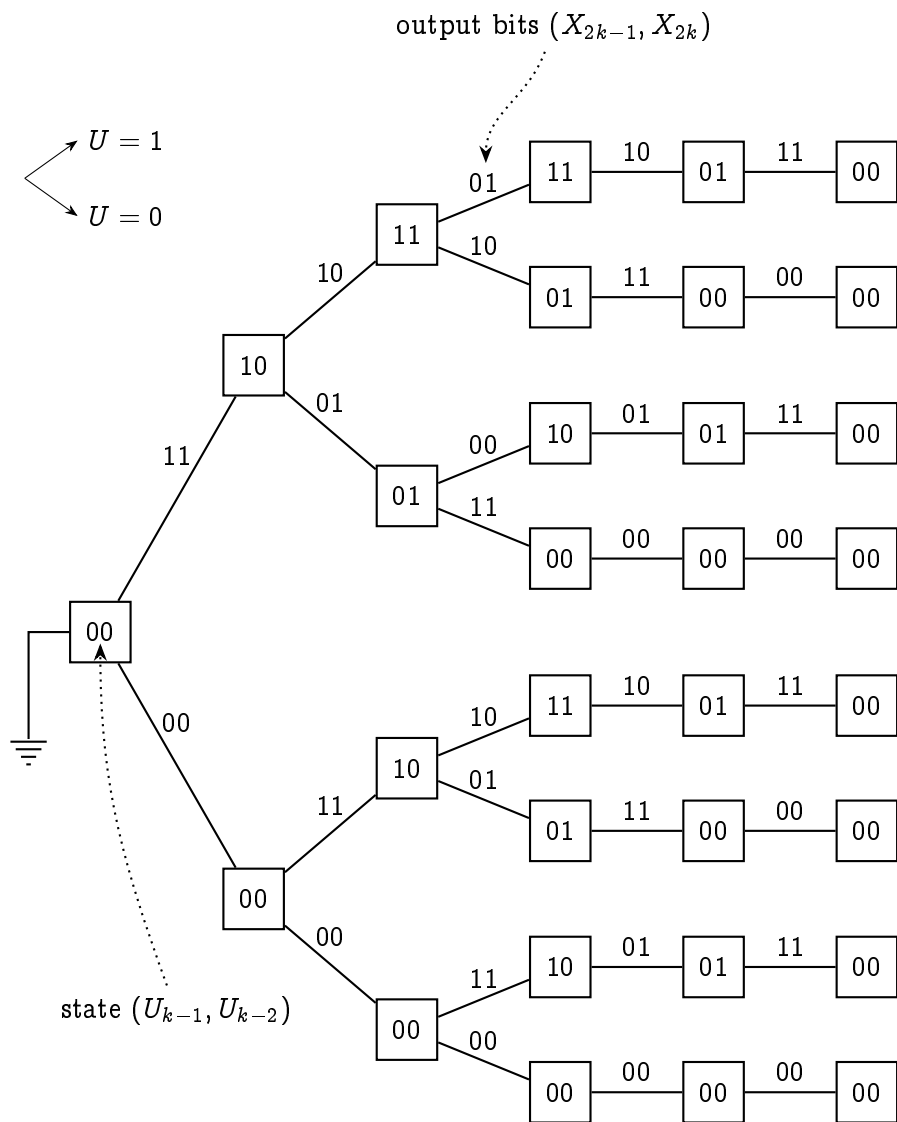


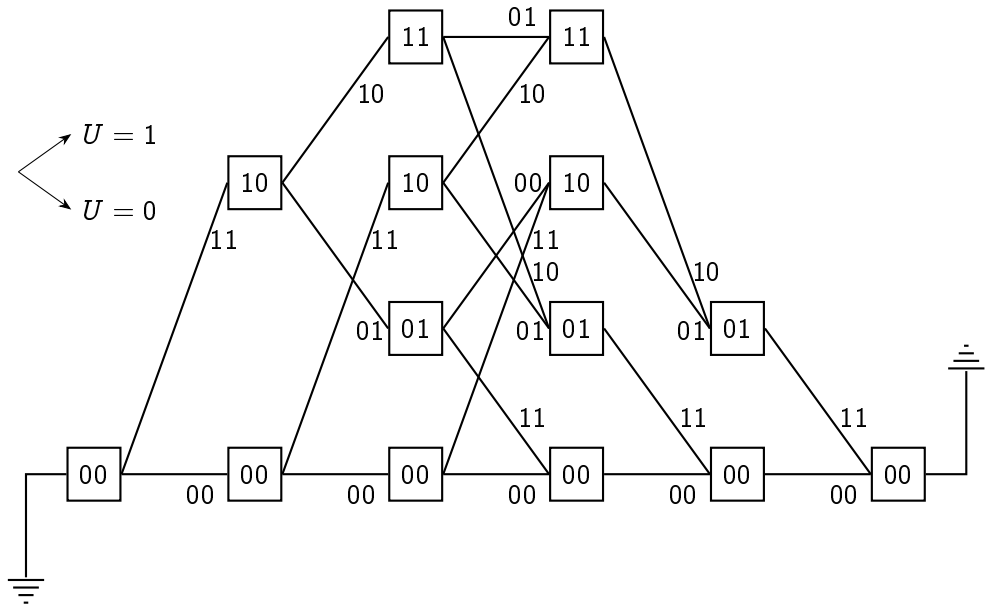


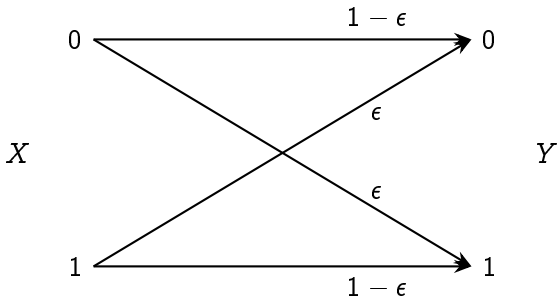
W

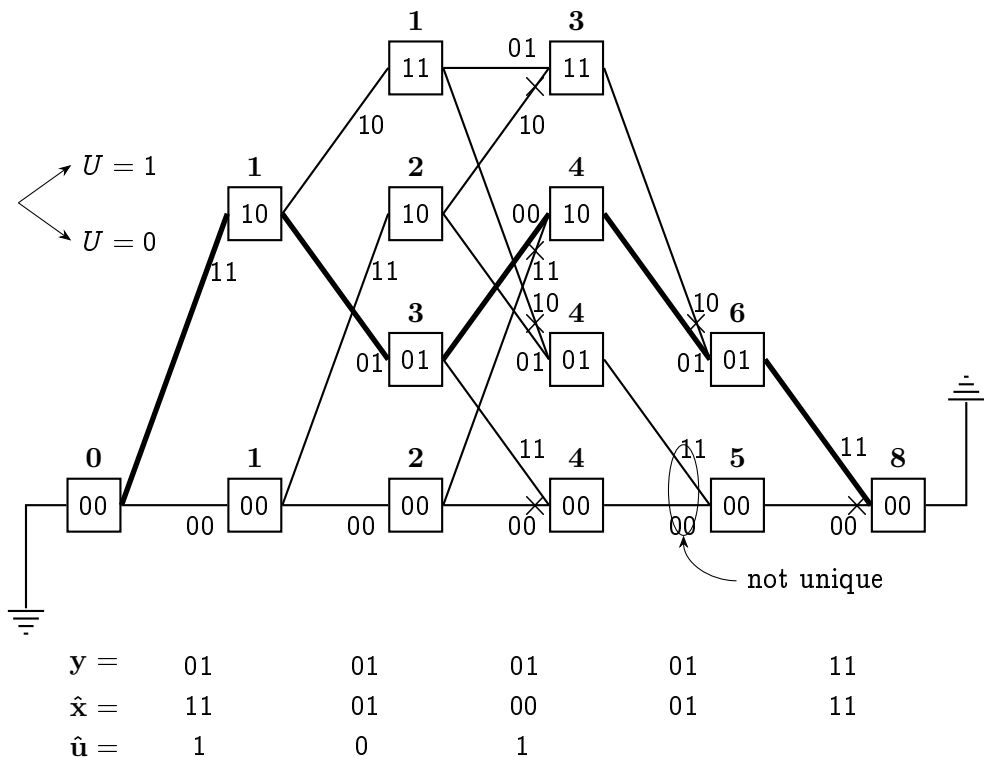


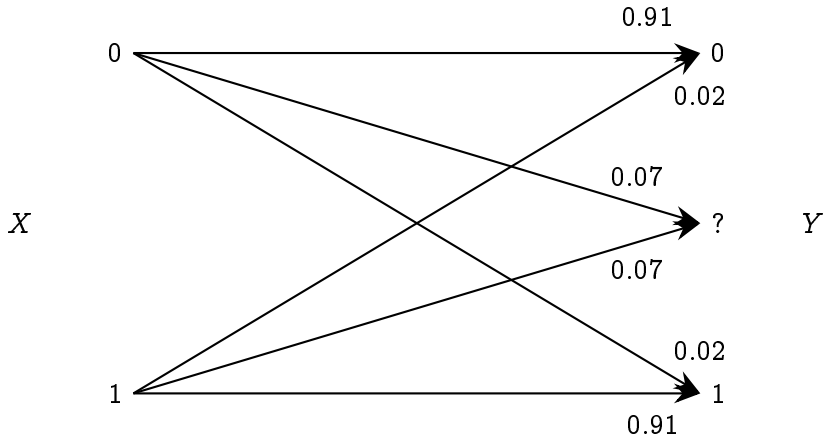






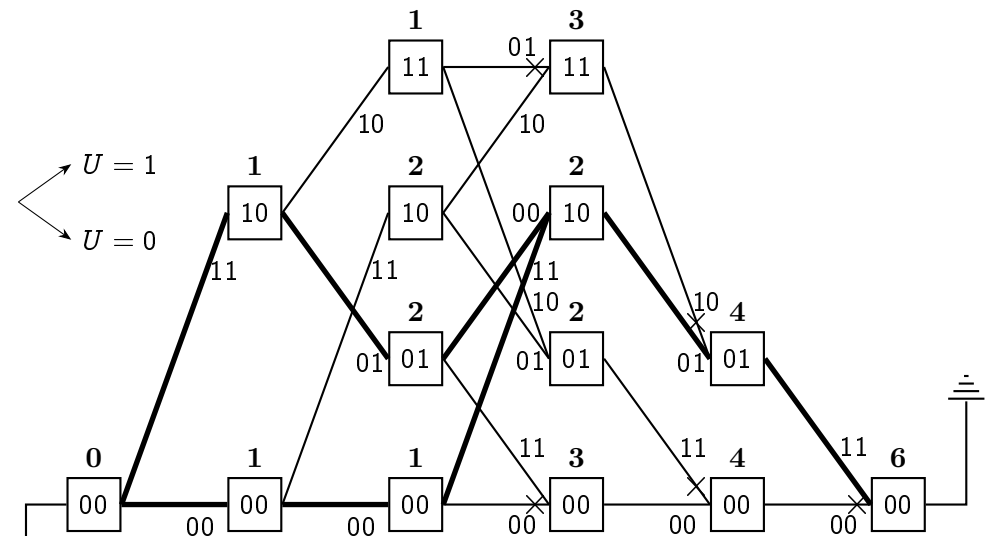




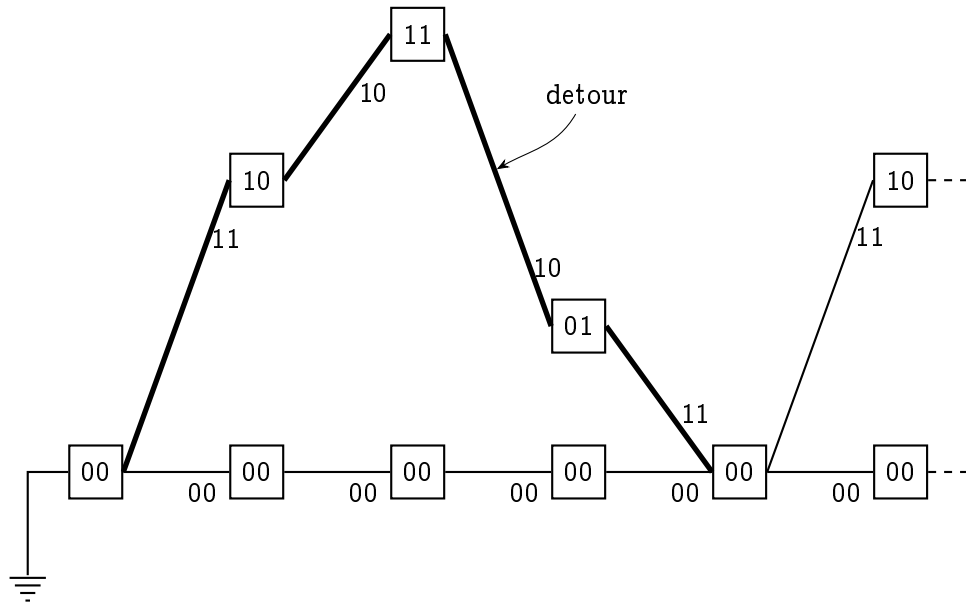


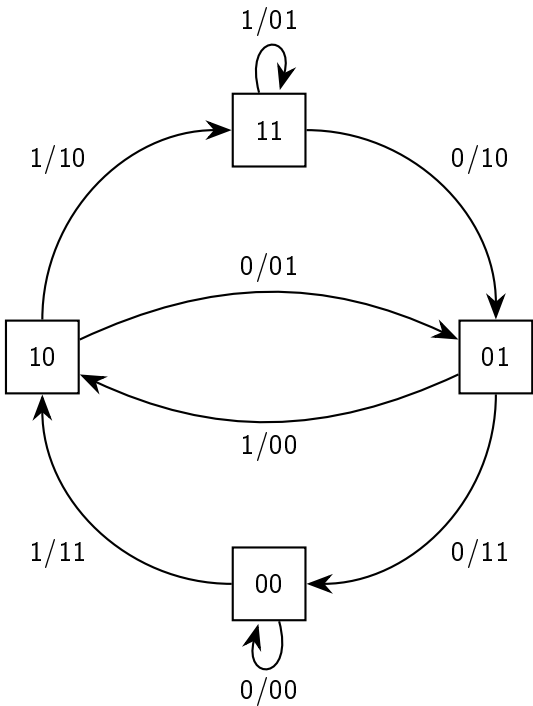
	$y = 0$	$y = ?$	$y = 1$
$x = 0$	$\log 0.91 - \log 0.02$ ≈ 5.50	$\log 0.07 - \log 0.07$ $= 0$	$\log 0.02 - \log 0.02$ $= 0$
$x = 1$	$\log 0.02 - \log 0.02$ $= 0$	$\log 0.07 - \log 0.07$ $= 0$	$\log 0.91 - \log 0.02$ ≈ 5.50

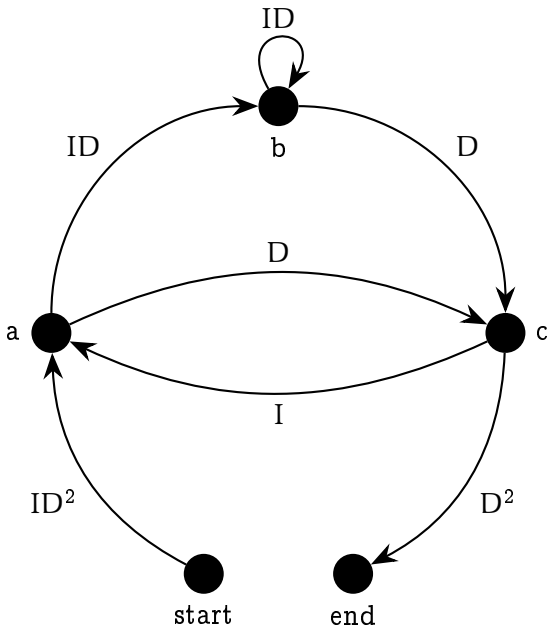
	$y = 0$	$y = ?$	$y = 1$
$x = 0$	1	0	0
$x = 1$	0	0	1

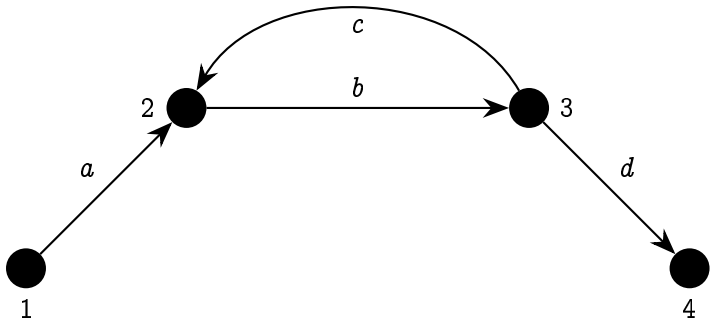


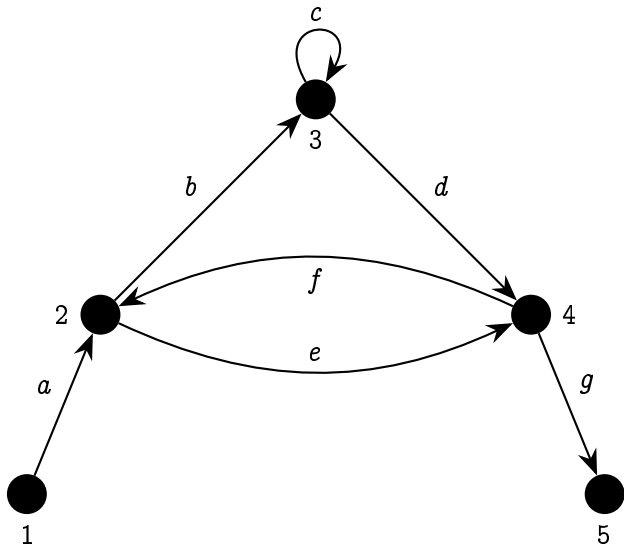
$\mathbf{y} =$	01	?1	1?	01	11
$\hat{\mathbf{x}} =$	11	01	00	01	11
	00	00	11	01	11
$\hat{\mathbf{u}} =$	1	0	1		
	0	0	1		

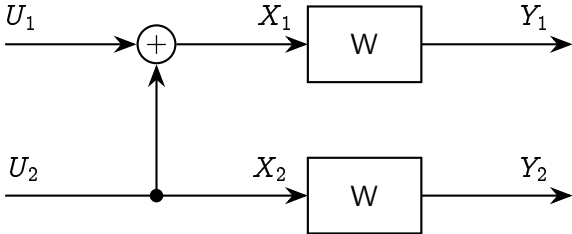


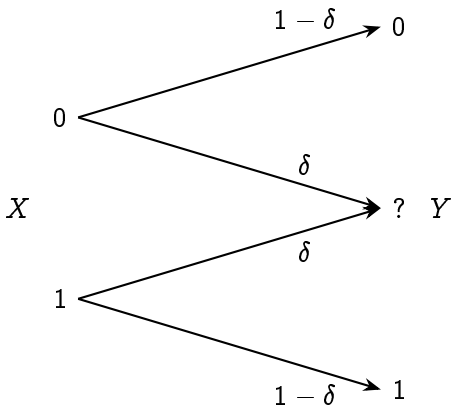




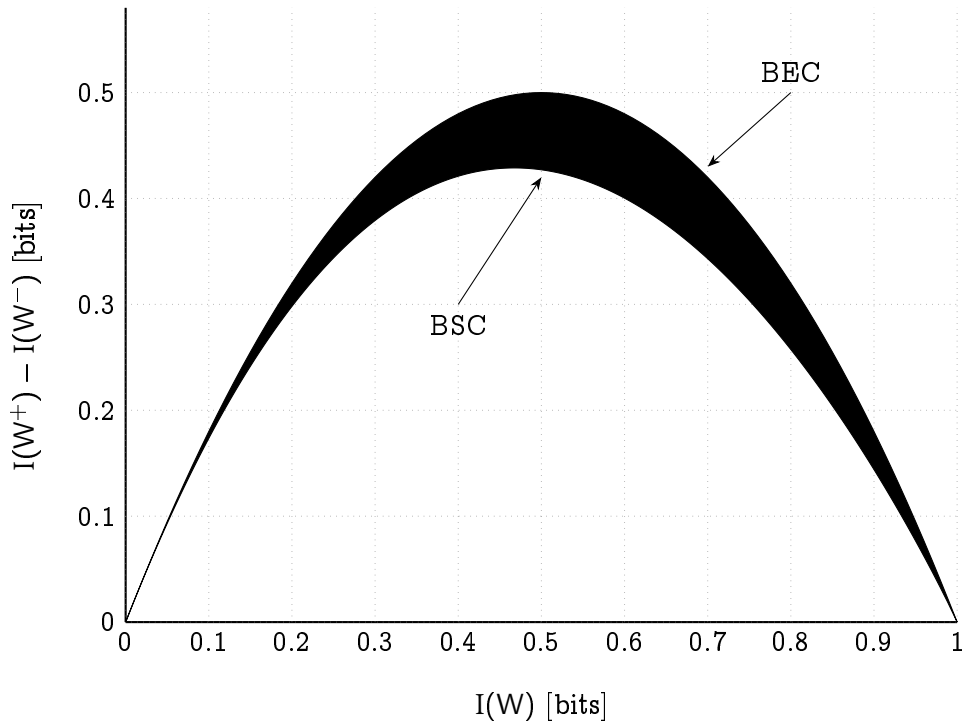


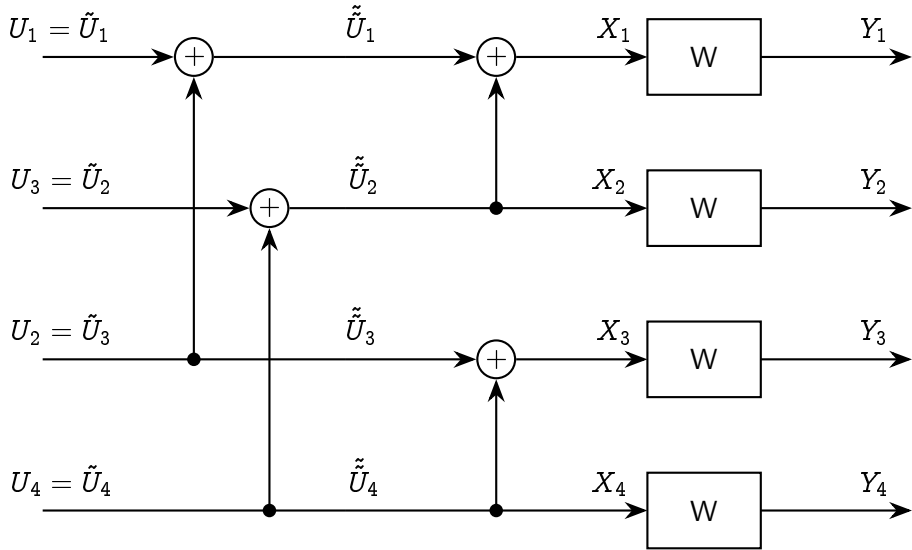


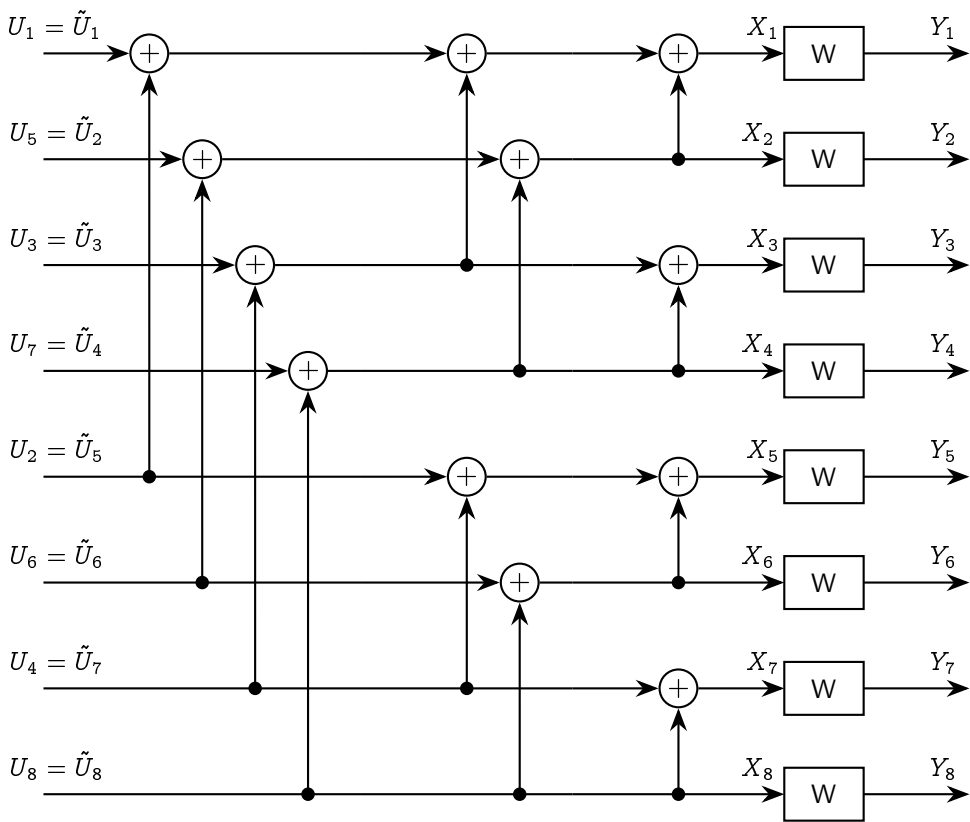


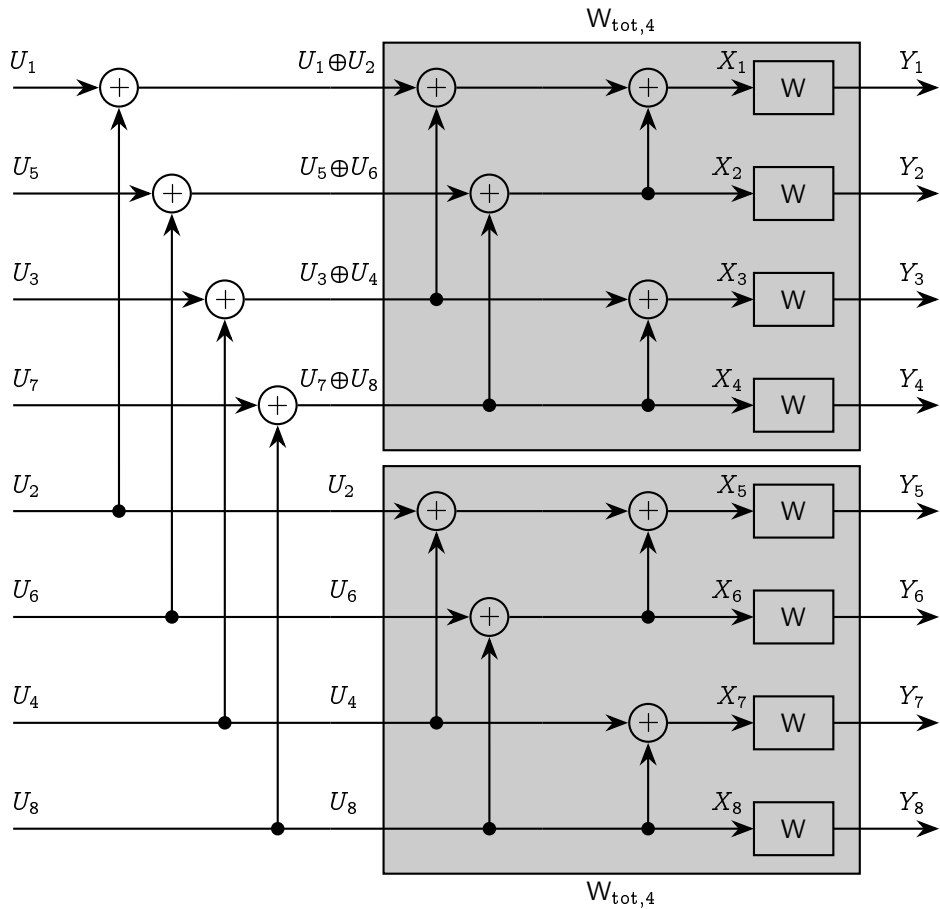


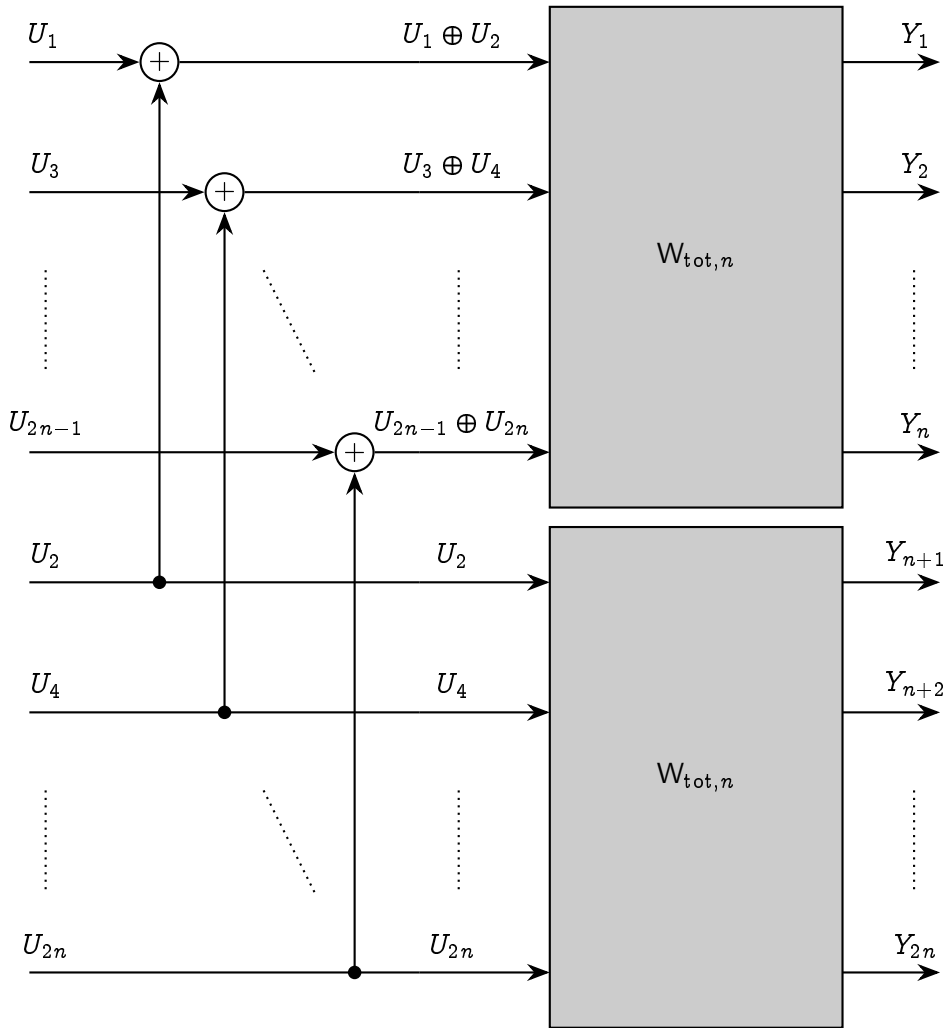
y_1	y_2	$W^-(y_1, y_2 0)$	$W^-(y_1, y_2 1)$
0	0	$(1 - \delta)^2/2$	0
0	?	$\delta(1 - \delta)/2$	$\delta(1 - \delta)/2$
0	1	0	$(1 - \delta)^2/2$
?	0	$\delta(1 - \delta)/2$	$\delta(1 - \delta)/2$
?	?	δ^2	δ^2
?	1	$\delta(1 - \delta)/2$	$\delta(1 - \delta)/2$
1	0	0	$(1 - \delta)^2/2$
1	?	$\delta(1 - \delta)/2$	$\delta(1 - \delta)/2$
1	1	$(1 - \delta)^2/2$	0

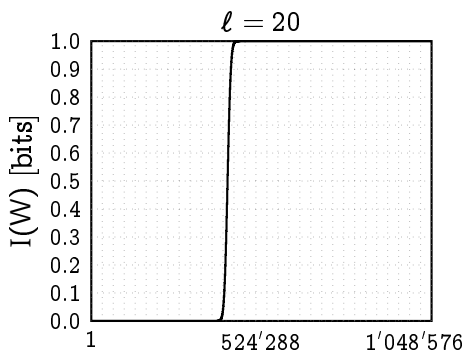
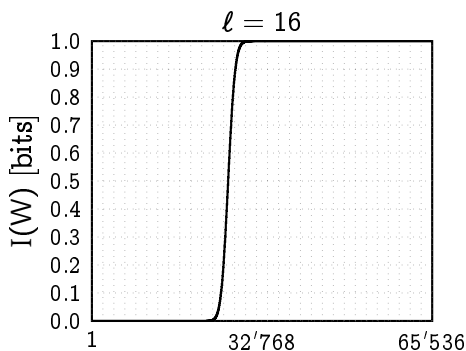
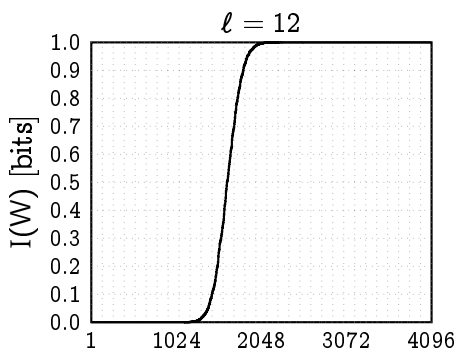
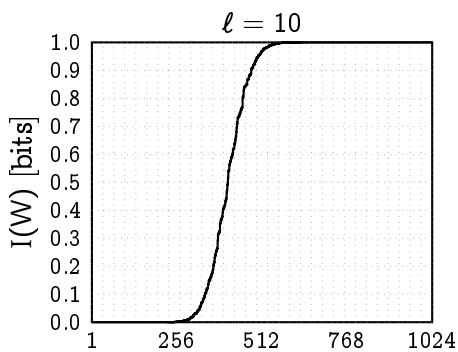
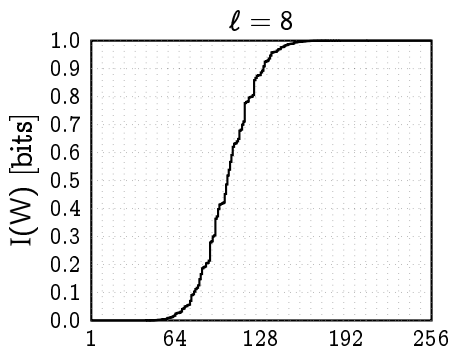
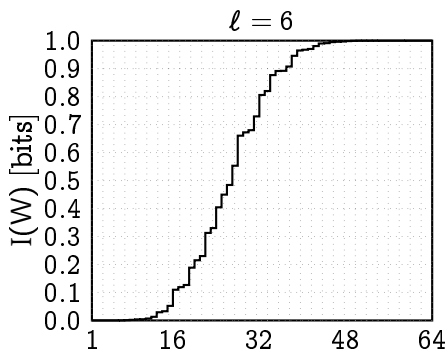
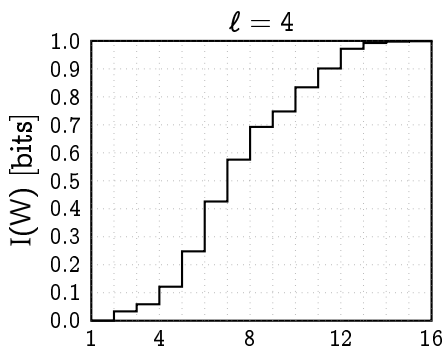
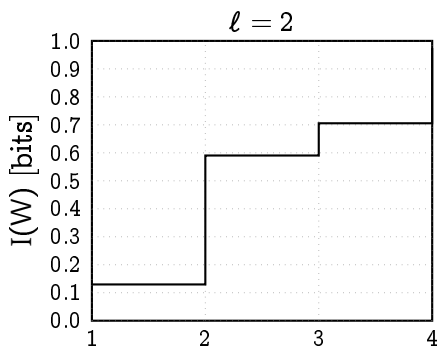


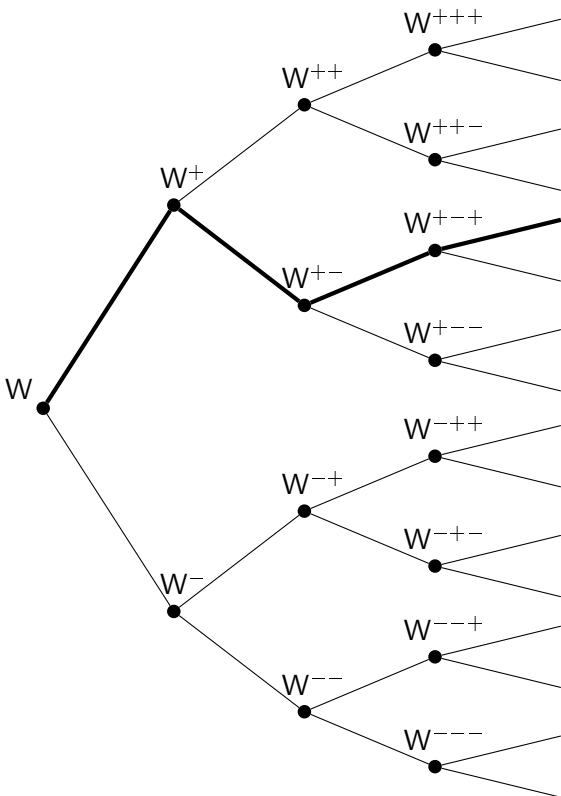


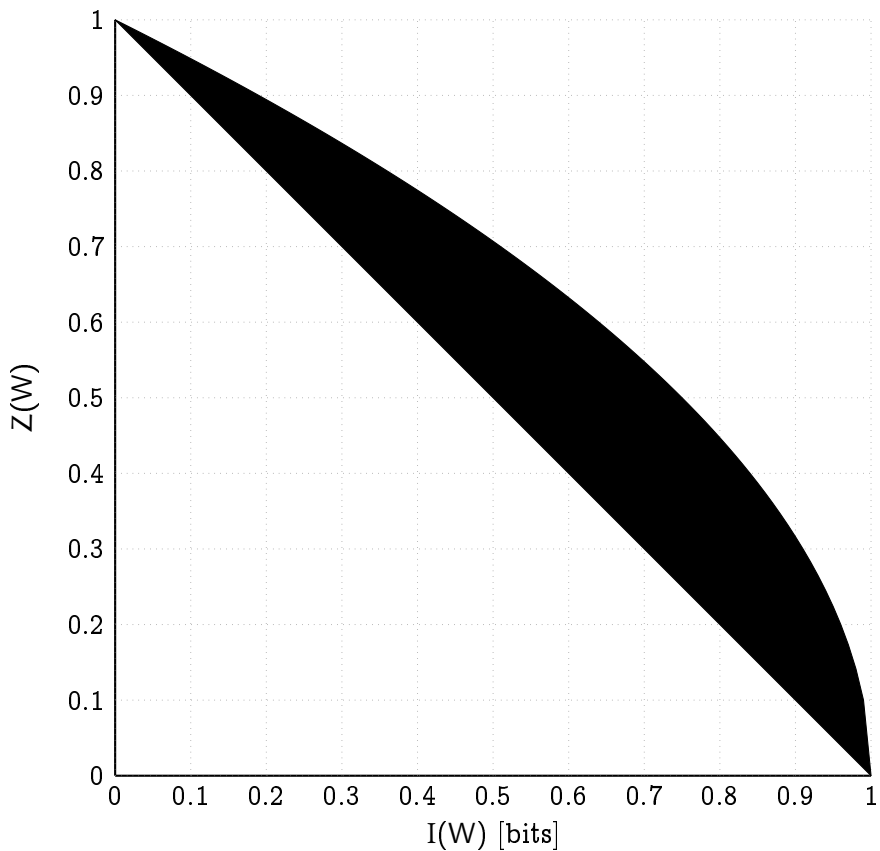


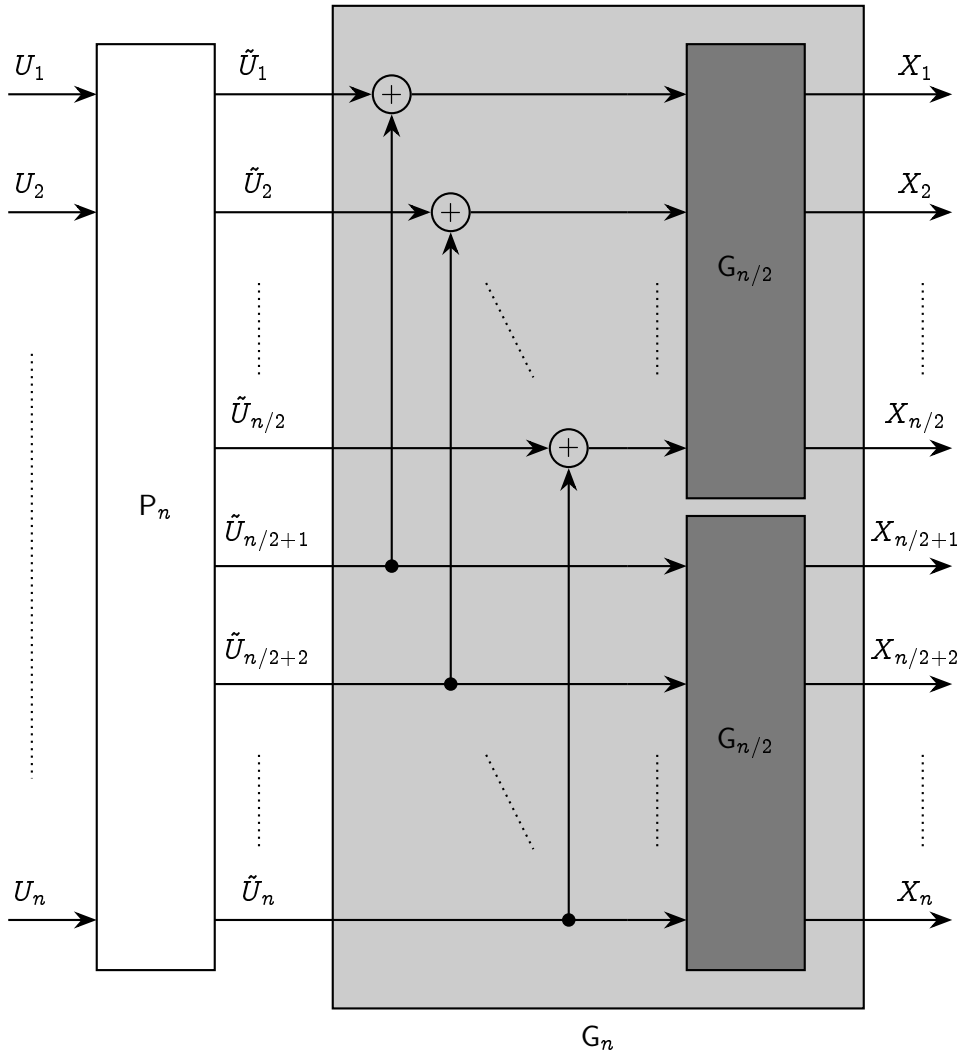


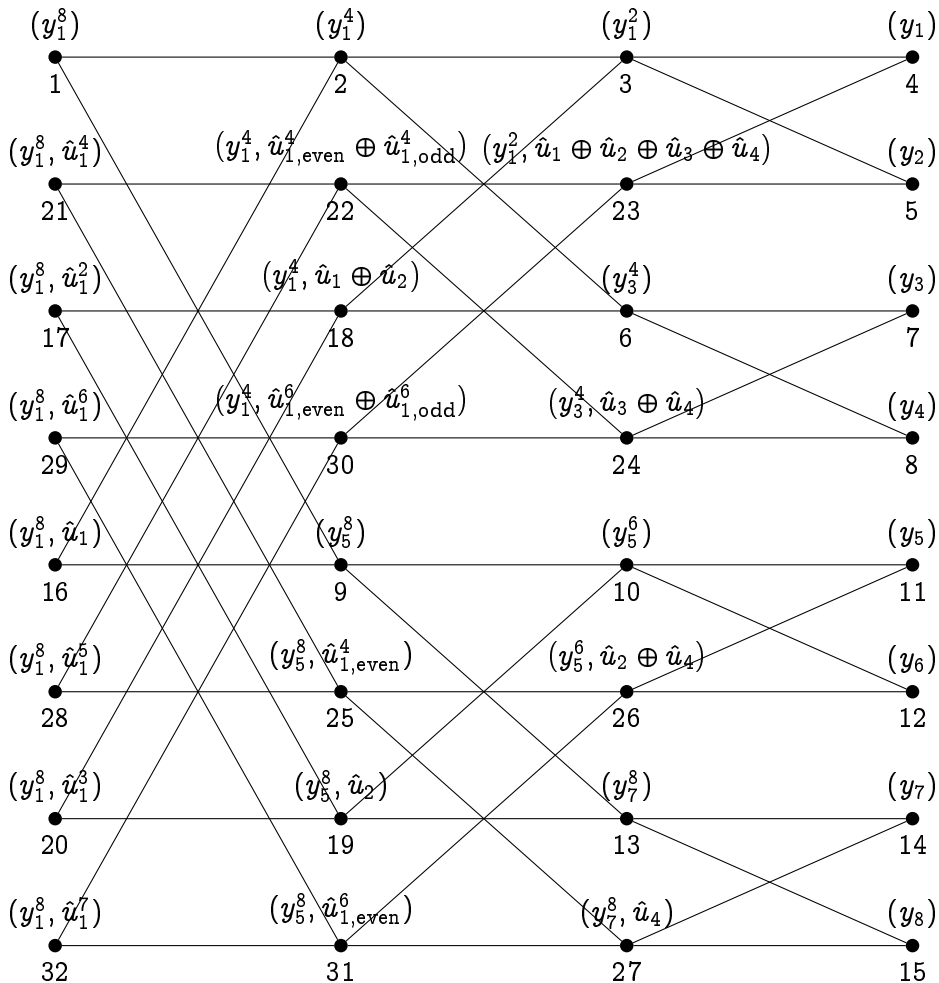




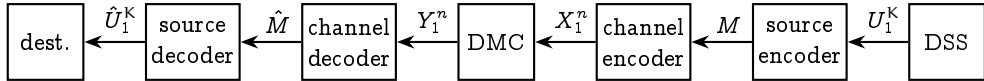


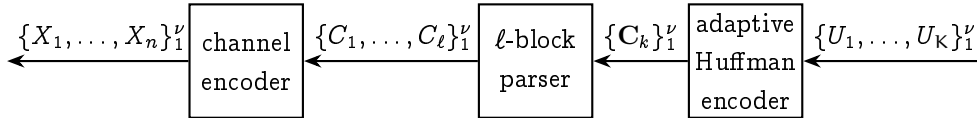


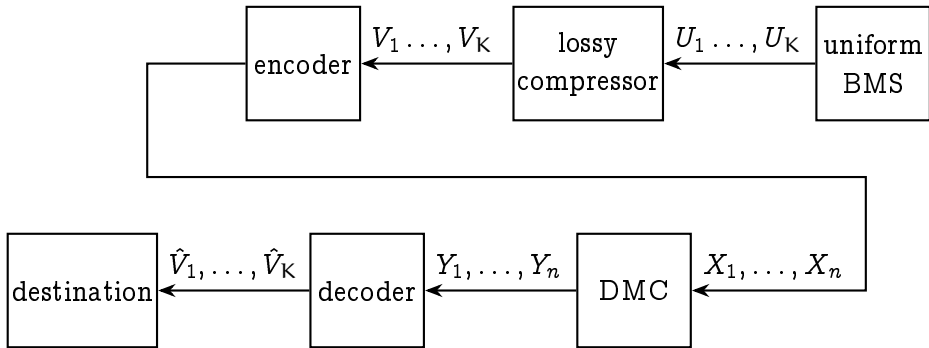


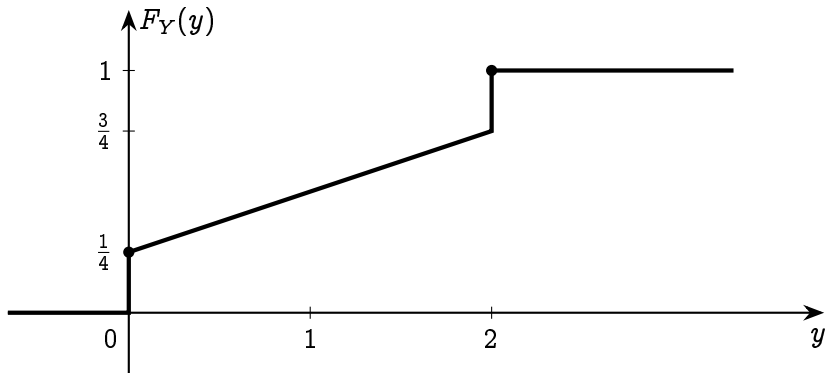


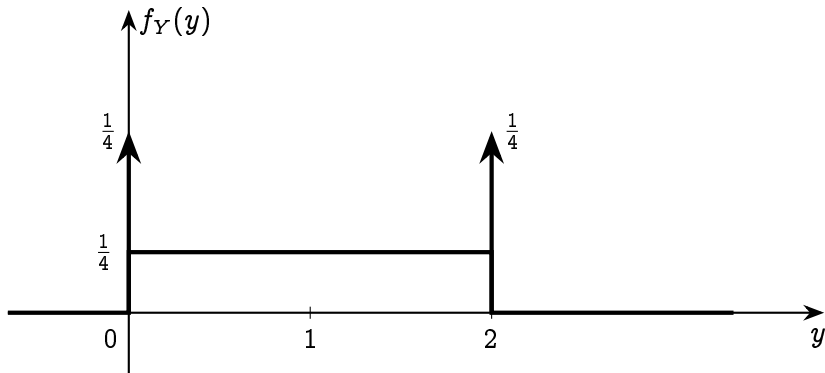


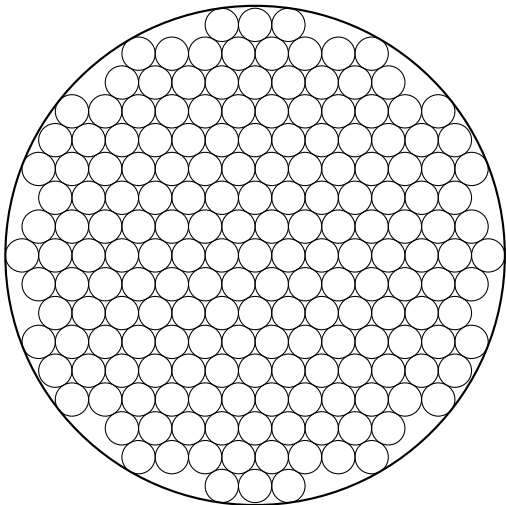


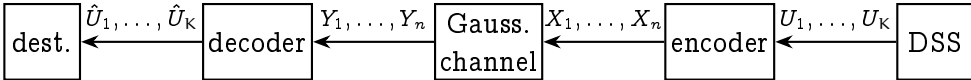


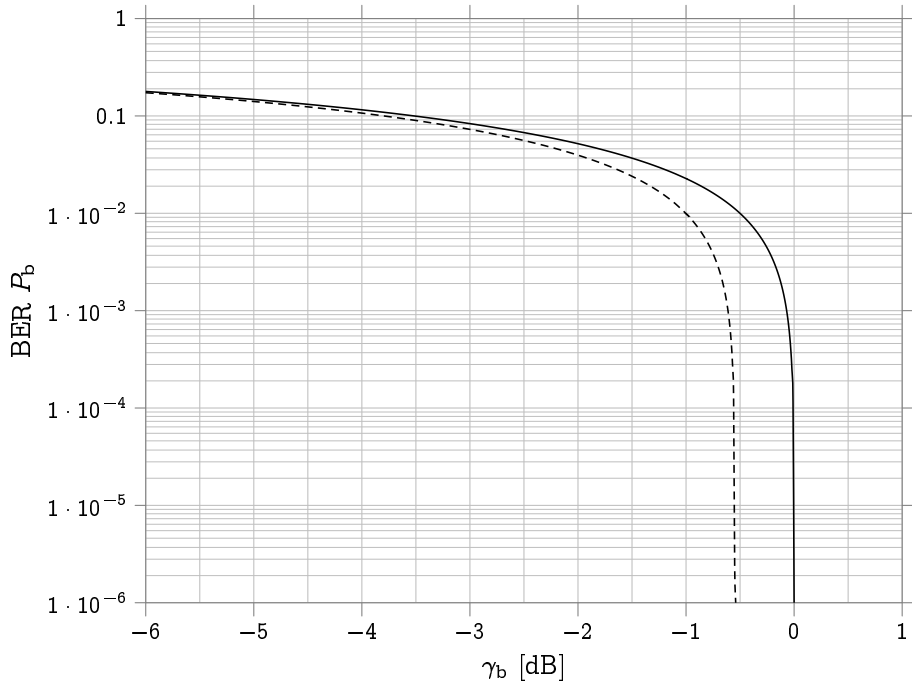




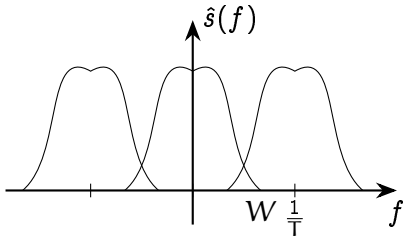
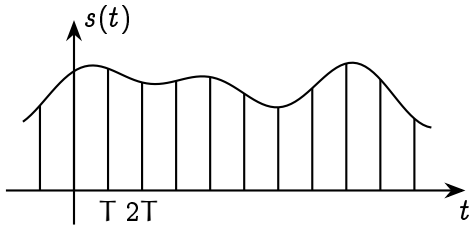


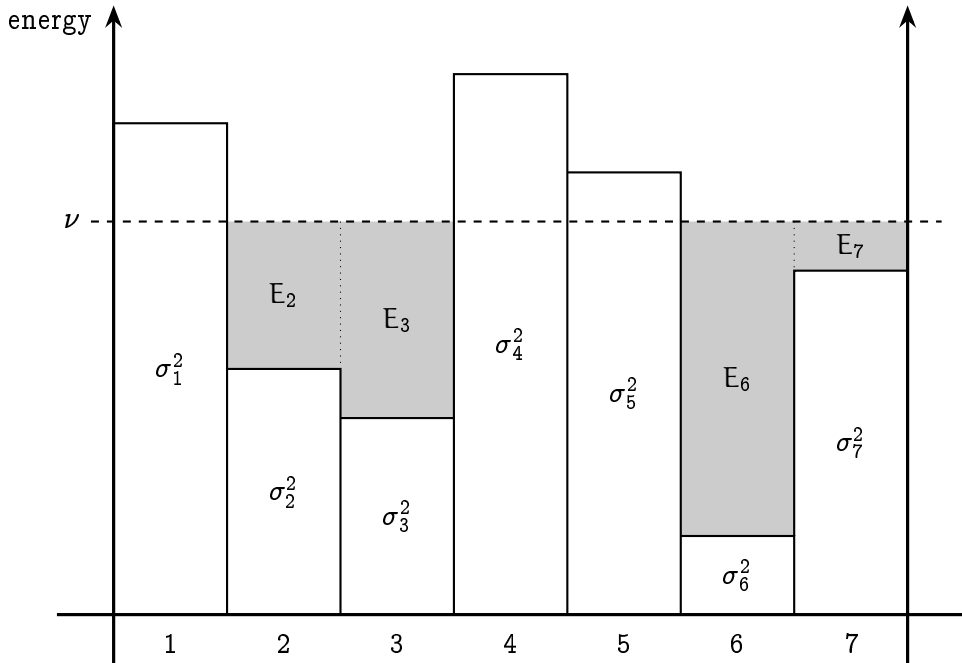


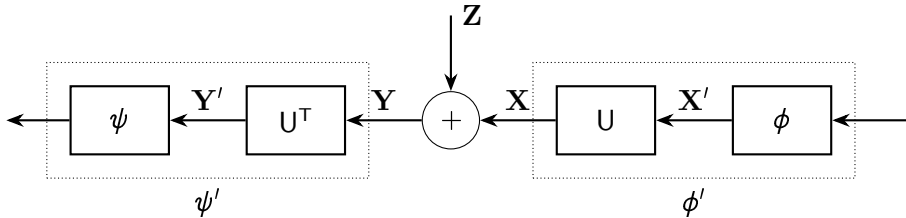


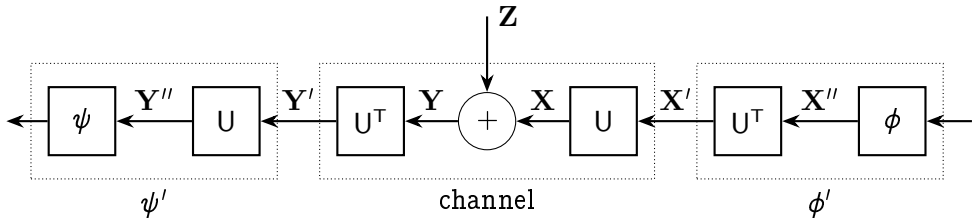


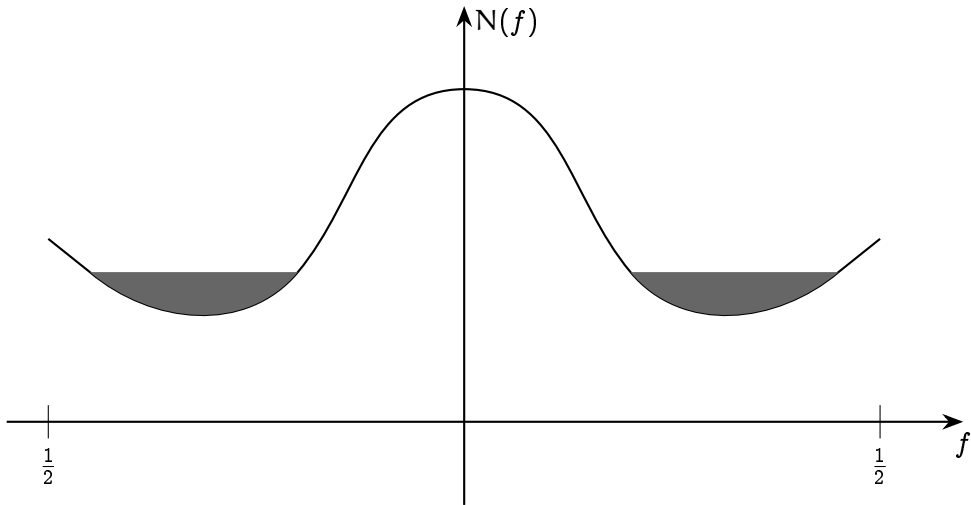
— $R = 1/2$ - - - $R = 1/3$



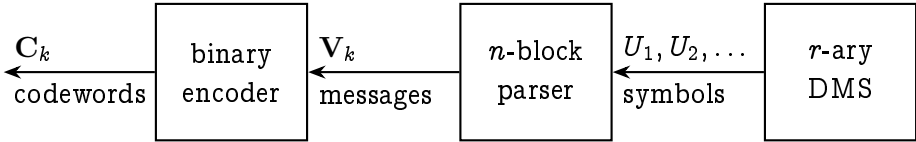








sequence	probability	
1111111111	0.7^{10}	
1111111110	$0.7^9 \cdot 0.3^1$	
1111111101	$0.7^9 \cdot 0.3^1$	
$\vdots$	$\vdots$	
0111111111	$0.7^9 \cdot 0.3^1$	
1111111100	$0.7^8 \cdot 0.3^2$	
$\vdots$	$\vdots$	
0011111111	$0.7^8 \cdot 0.3^2$	
1111111000	$0.7^7 \cdot 0.3^3 = 2^{10} H(X)$	typical
$\vdots$	$\vdots$	$\vdots$
0001111111	$0.7^7 \cdot 0.3^3 = 2^{10} H(X)$	typical
1111110000	$0.7^6 \cdot 0.3^3$	
$\vdots$	$\vdots$	
0000111111	$0.7^6 \cdot 0.3^3$	
$\vdots$	$\vdots$	
0000000000	0.3^{10}	



$P_{X,Y}$	$Y = 0$	$Y = 1$	$Y = 2$	P_X
$X = 0$	$\frac{1}{2}$	0	$\frac{1}{4}$	$\frac{3}{4}$
$X = 1$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
P_Y	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	

