



Tokyo Tech

マーカーを用いた挿入／削除／反転誤り 訂正符号化法

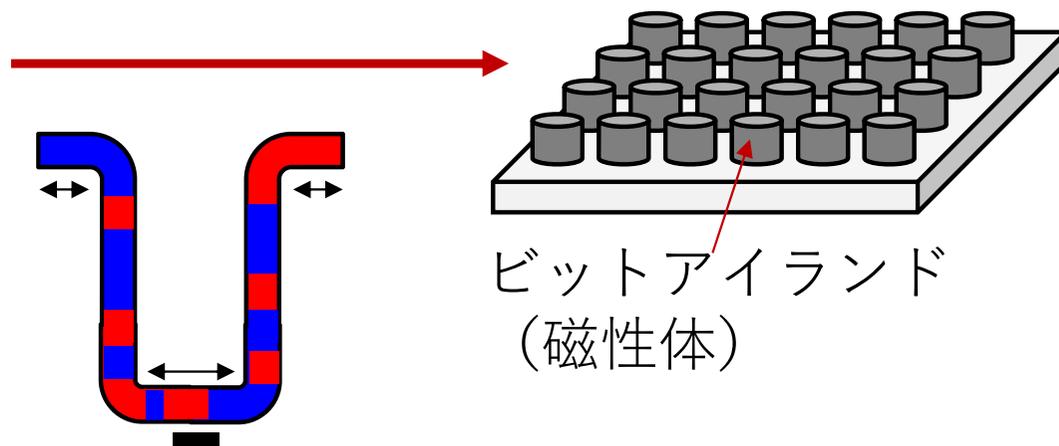
金子 晴彦

東京工業大学
情報理工学院

研究の背景

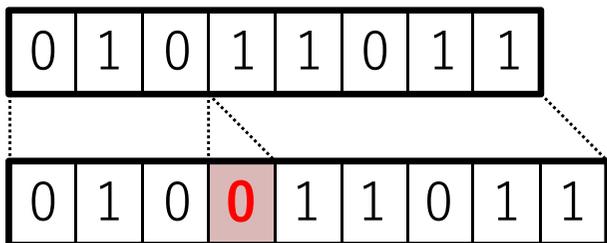
- 高密度ストレージ／メモリ
 - bit-patterned media (BPM)
 - racetrack memory
- 高速データバス
- DNAストレージ など

➡ 反転誤り + **挿入／削除誤り**

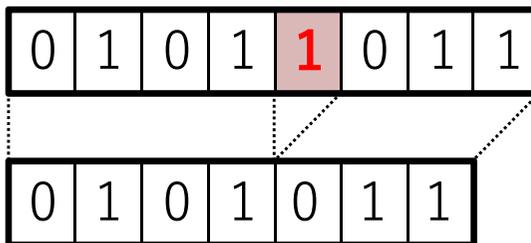


挿入／削除／反転(IDS)誤り

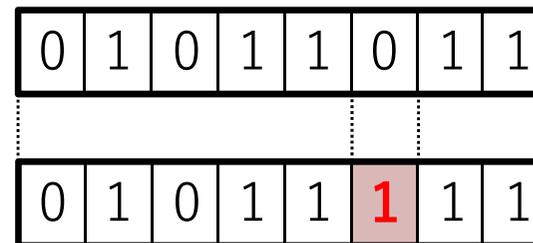
挿入 (Insertion)



削除 (Deletion)



反転 (Substitution)

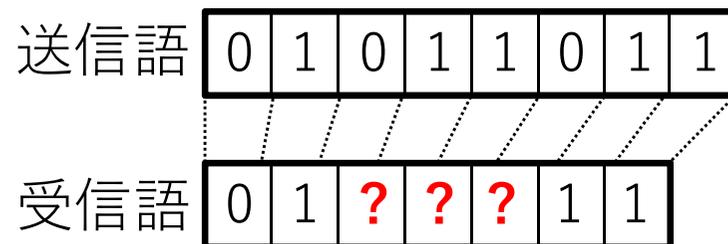
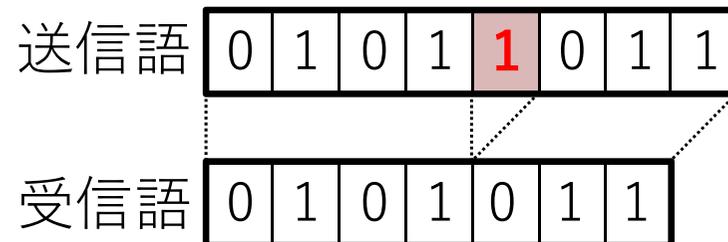


研究の目的：

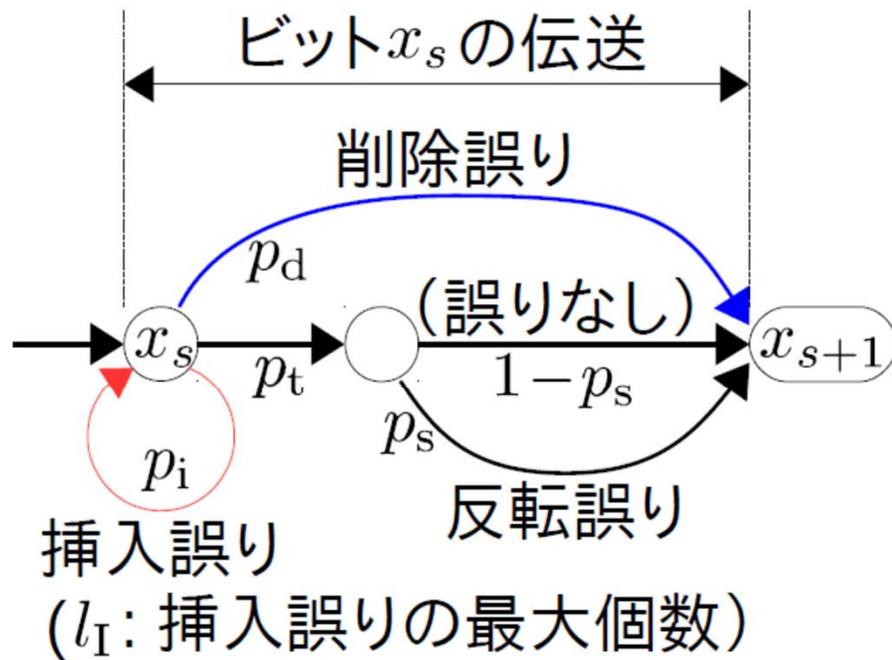
- 高信頼なストレージ，メモリ，データバス
 - IDS誤り通信路のモデル化
 - IDS誤り制御符号の構成・復号法

発表内容

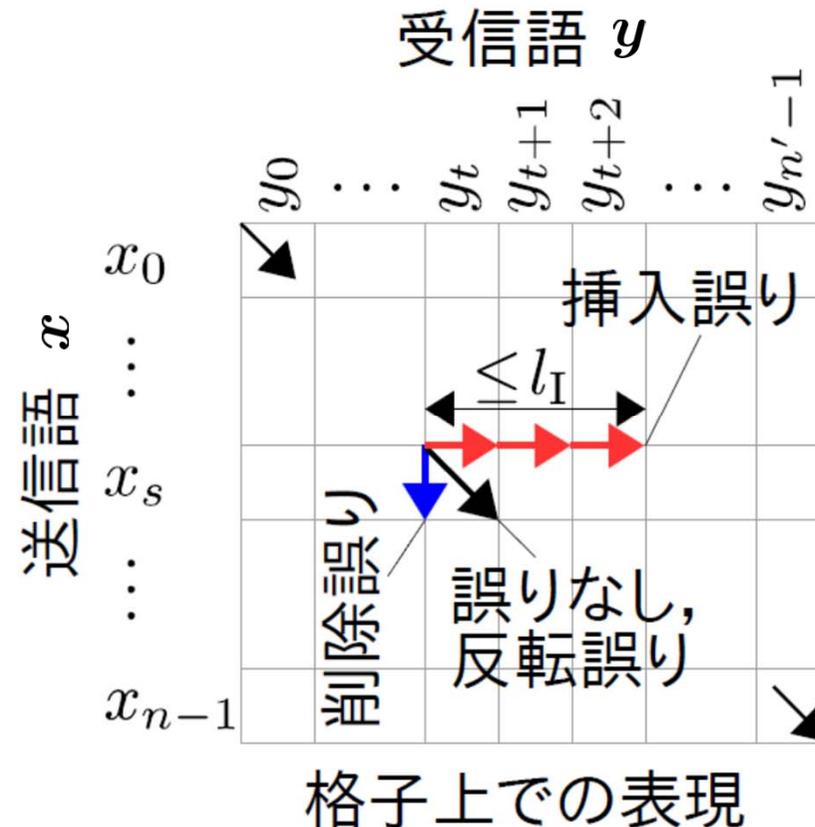
- ◆ 研究の背景
- ◆ 挿入／削除／反転(IDS)誤り通信路
 - ◆ 通信路モデル
 - ◆ 従来のIDS誤り訂正符号の種類
 - ◆ マーカーを用いた符号化と復号
- ◆ ドリフト通信路
 - ◆ 通信路モデル
 - ◆ マーカーを用いた符号化と復号
- ◆ まとめ



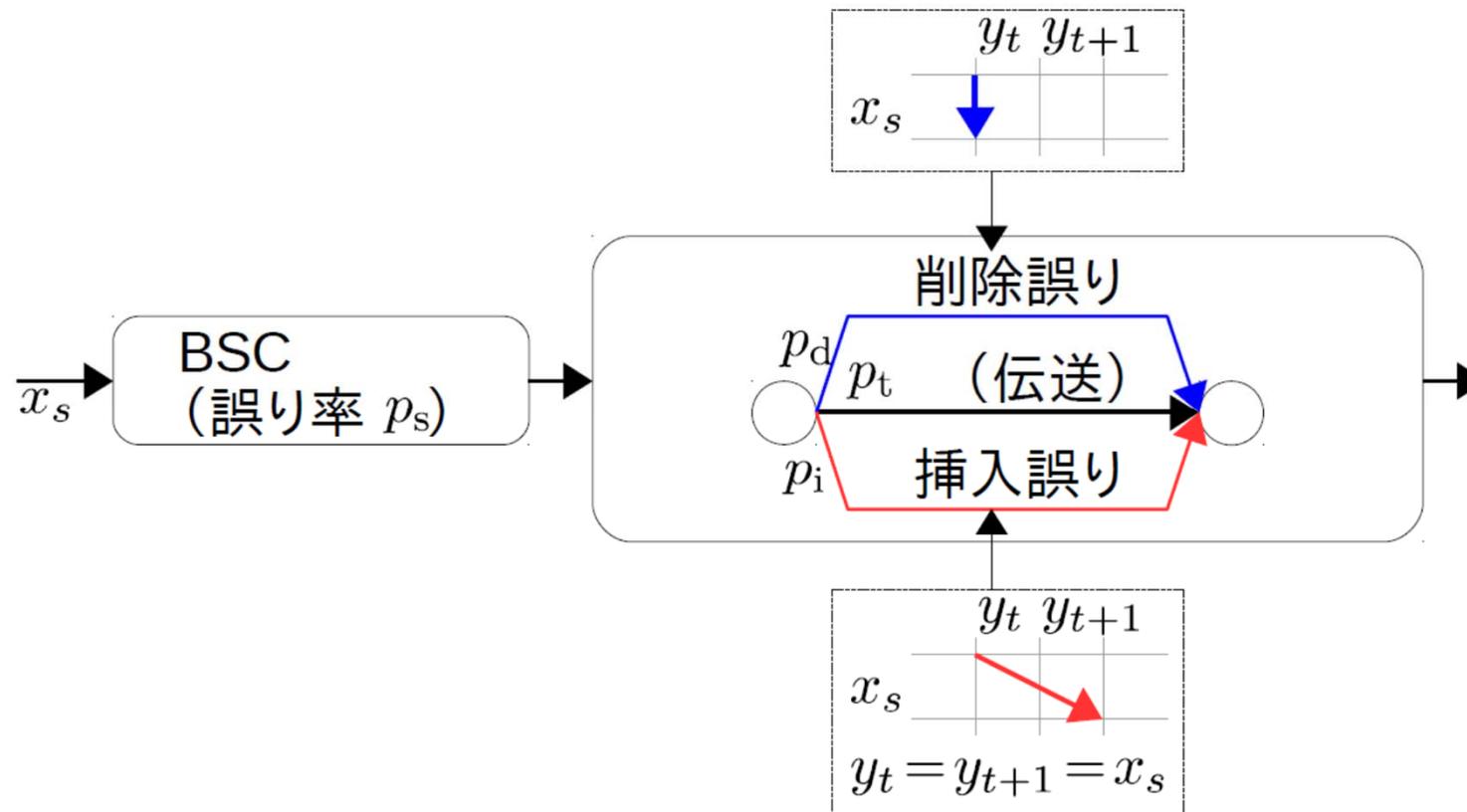
通信路モデルA



状態遷移図

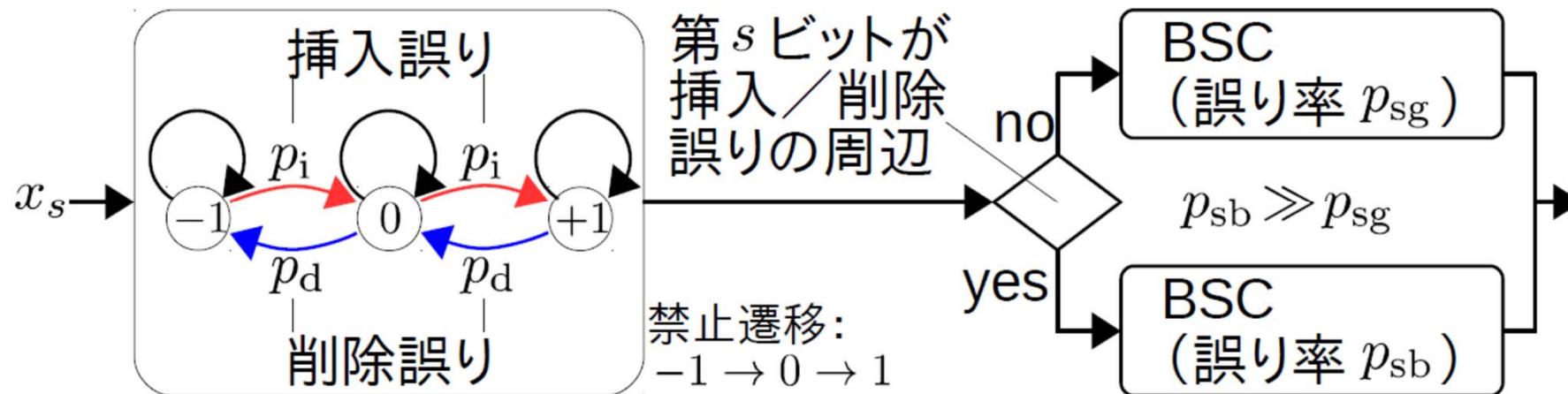


通信路モデルB



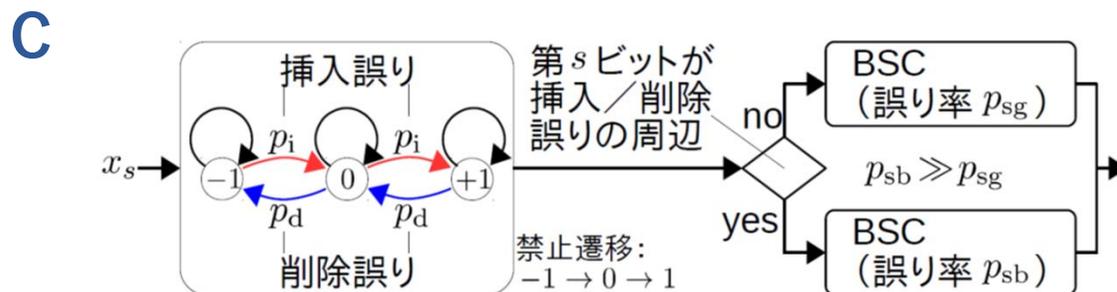
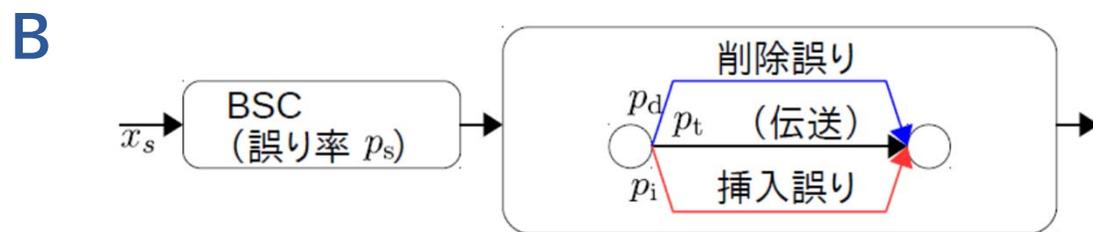
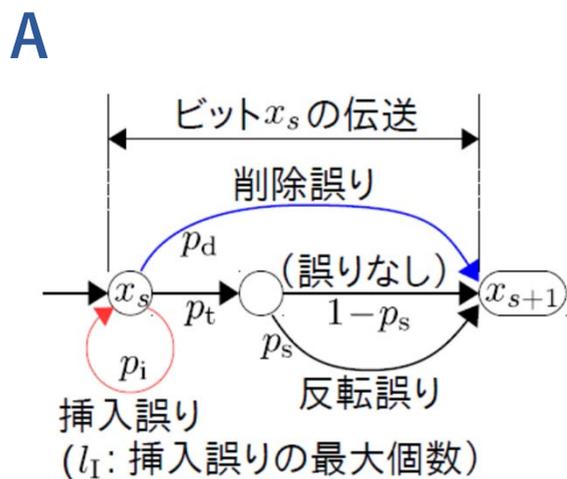
G. Han, Y. L. Guan, K. Cai, K. S. Chan, and L. Kong,
 "Coding and Detection for Channels With Written-In Errors and Inter-Symbol Interference,"
IEEE Trans. Magnetics, Vol. 50, No. 10, 3101106, Oct. 2014.

通信路モデルC



通信路モデルの比較

通信路モデル	A	B	C
送信ビット x_s における最大挿入ビット数	l_I	1	1
送信ビット x_s における挿入ビット e の値	$p(e=0) =$ $p(e=1) = \frac{1}{2}$	$e = x_s$	$e = x_s$
反転誤り確率	p_s	p_s	p_{sb} OR p_{sg} ($p_{sb} \gg p_{sg}$)



発表内容

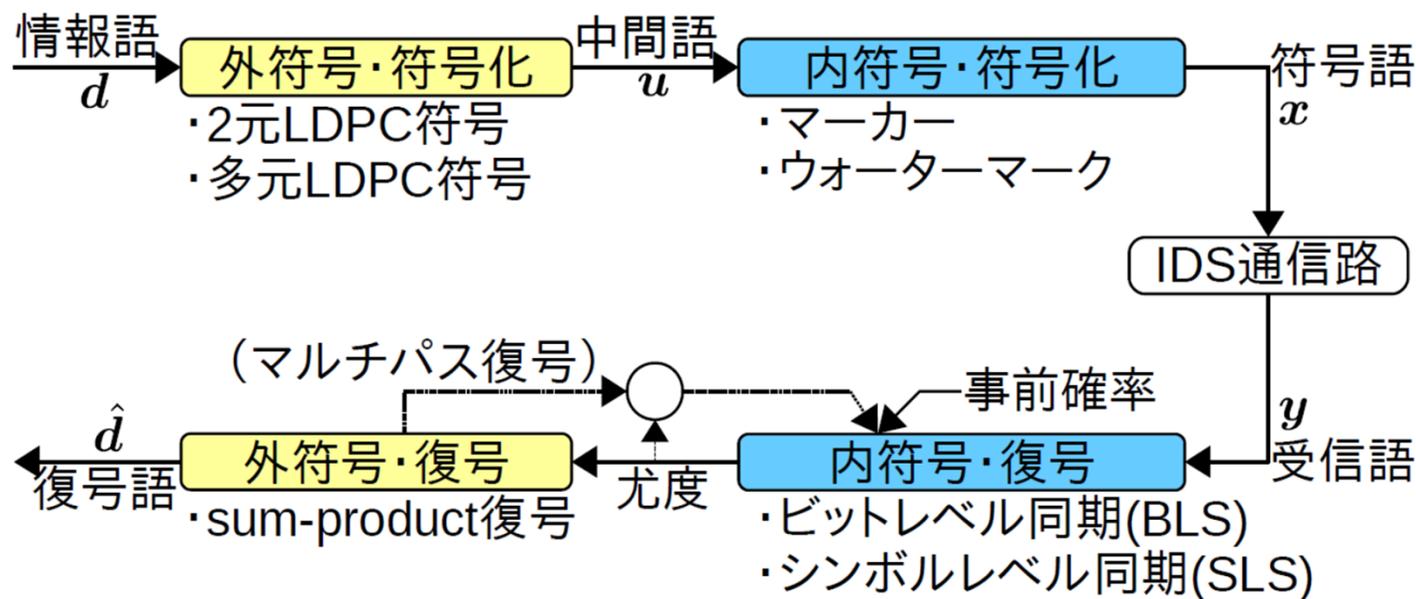
- ◆ 研究の背景
- ◆ 挿入／削除／反転(IDS)誤り通信路
 - ◆ 通信路モデル
 - ◆ 従来のIDS誤り訂正符号の種類
 - ◆ マーカーを用いた符号化と復号
- ◆ ドリフト通信路
 - ◆ 通信路モデル
 - ◆ マーカーを用いた符号化と復号
- ◆ まとめ

IDS誤り訂正符号の例

1. 数論的な構成による符号

$$C = \left\{ \mathbf{u} = (u_0, u_1, \dots, u_{n-1}) \mid \sum_{i=0}^{n-1} a_i u_i \bmod M = b \right\} \quad a_i: \text{係数}$$

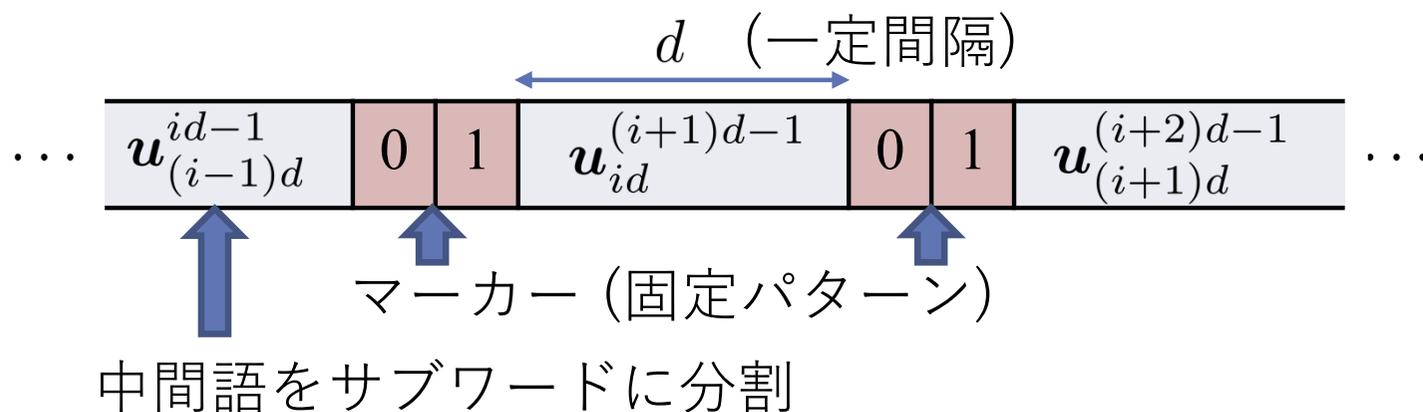
2. マーカーを用いた符号化／ウォーターマークを用いた符号化



3. 空間結合符号を用いた符号化

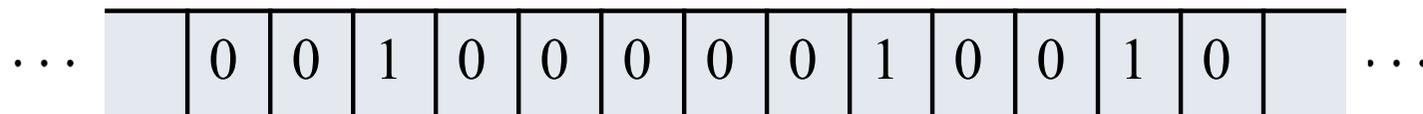
マーカー・ウォータマークによる同期

マーカーによる同期

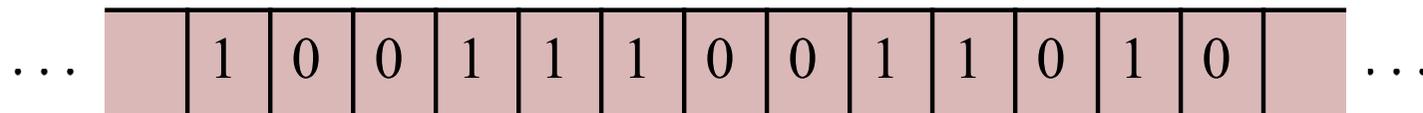


ウォータマーク (WM)による同期

中間語を疎なベクトルに変換



\oplus



ランダムな系列 (復号器と共有)

マーカー・ウォータマークを用いた符号の例

文献	[4] Davey 2001	[5] Feng 2011	[6] Wang 2011	[7] Kaneko 2014
通信路	IDS	IDS	削除のみ	IDS
内符号	WM		マーカー	
内符号の復号	BLS		SLS	
シンボルサイズ b	1	≤ 3	8, 12, 16	≤ 8
外符号(LDPC符号)	多元	2元	2元	多元

送信語 : $\mathbf{x} = (x_0, x_1, \dots, x_{n-1}) \in \mathbb{B}^n$ ($\mathbb{B} = \{0, 1\}$)

受信語 : $\mathbf{y} = (y_0, y_1, \dots, y_{n'-1}) \in \mathbb{B}^{n'}$

Bit-level synchronization

$p(\mathbf{y} | x_s = a)$: 送信ビットごとの尤度を計算
($a \in \mathbb{B}$)

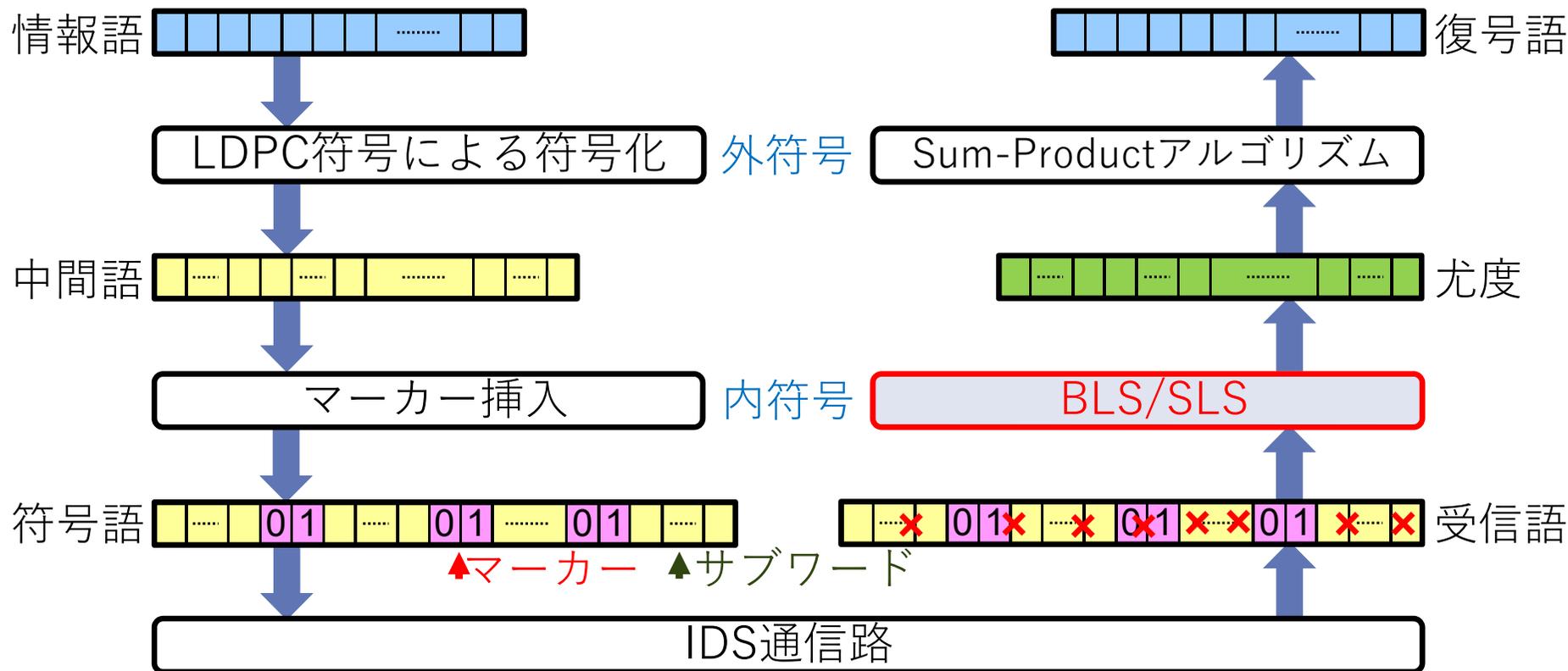
Symbol-level synchronization

$p(\mathbf{y} | \mathbf{x}_s^{s+b-1} = \mathbf{a})$: 送信シンボル(b ビット)ごとの尤度を計算
($\mathbf{a} \in \mathbb{B}^b$)

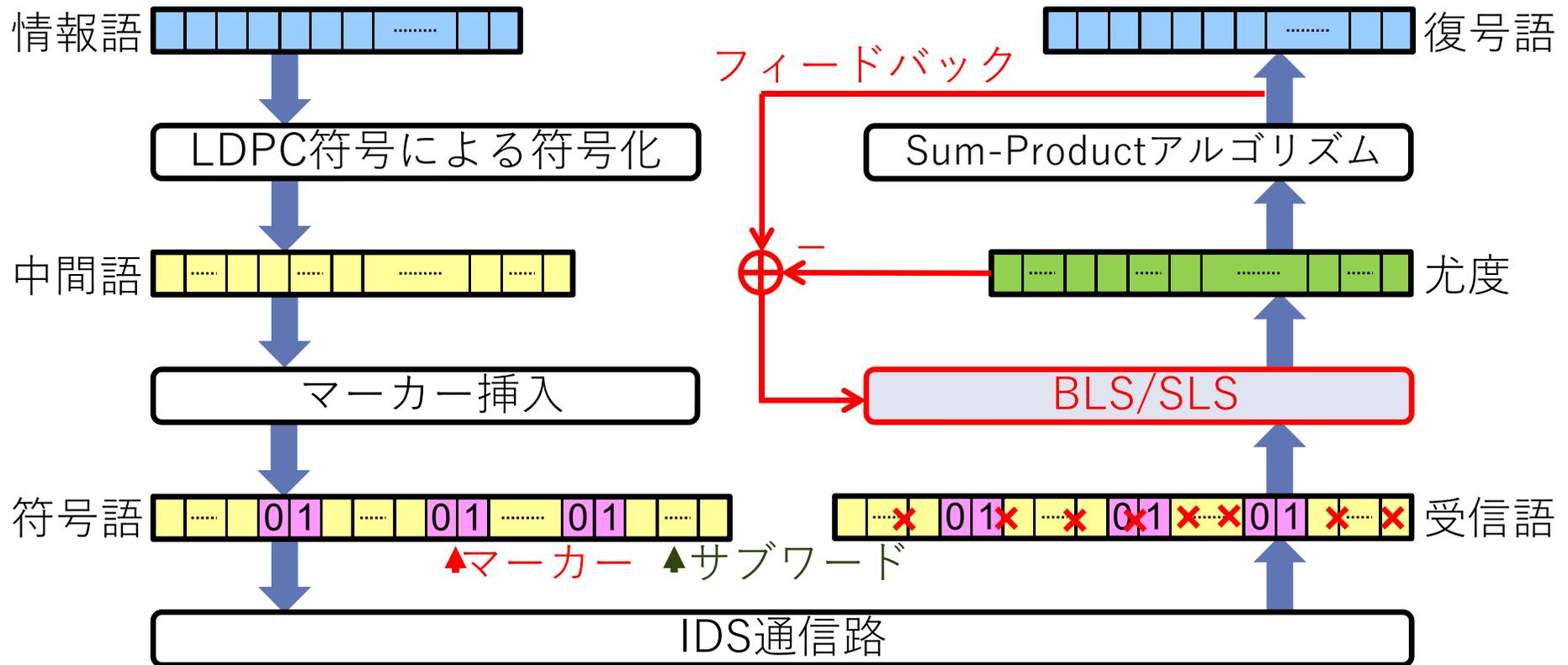
発表内容

- ◆ 研究の背景
- ◆ 挿入／削除／反転(IDS)誤り通信路
 - ◆ 通信路モデル
 - ◆ 従来のIDS誤り訂正符号の種類
 - ◆ マーカーを用いた符号化と復号
 - ◆ BLS/SLS
 - ◆ ルックアップテーブルによるSLSの高速化
- ◆ ドリフト通信路
 - ◆ 通信路モデル
 - ◆ マーカーを用いた符号化と復号
- ◆ まとめ

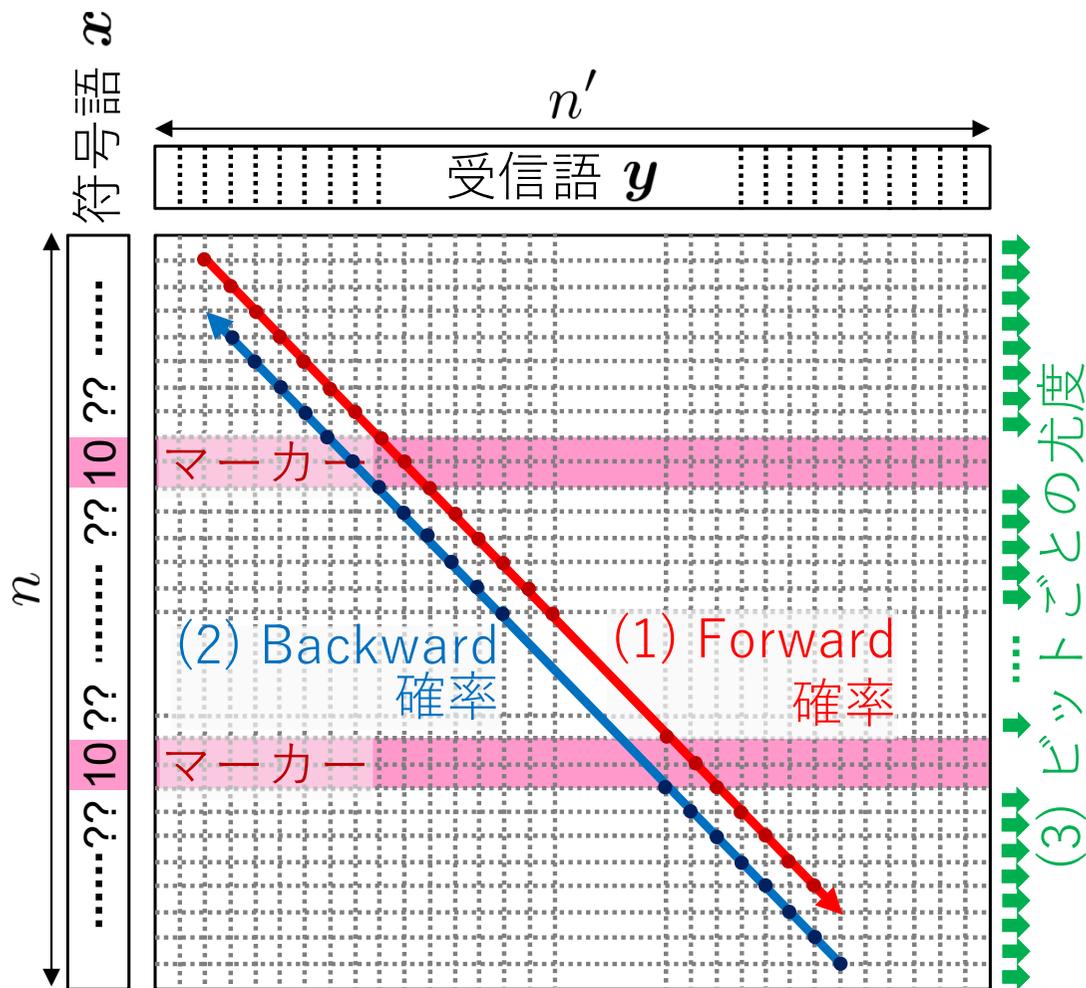
符号化・復号の流れ



マルチパス復号

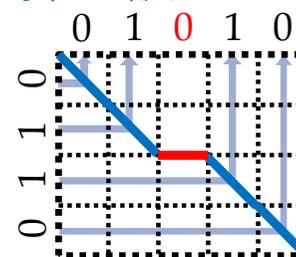


Bit-Level Synchronization (BLS)

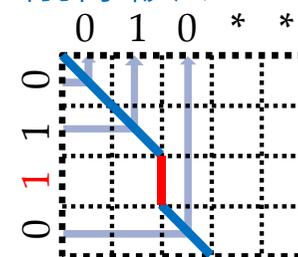


IDS通信路上での符号語の伝送：
 左上 $(0, 0)$ から右下 (n, n') に
 至るパスで表現

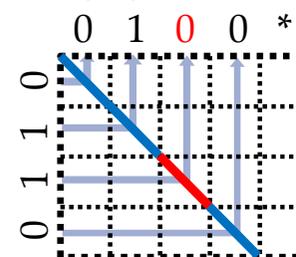
挿入誤り



削除誤り



反転誤り



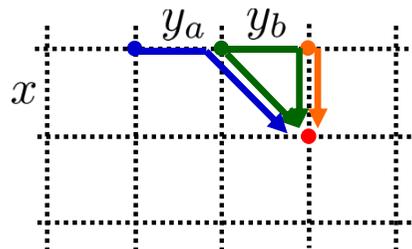
Forward 確率: $\alpha(s, t) = p(\mathbf{y}_0^{t-1}, E_{s,t})$

Backward 確率: $\beta(s, t) = p(\mathbf{y}_t^{n'-1} | E_{s,t})$

$E_{s,t}$: s ビット送信したとき
 t ビット受信している事象

BLSにおける計算

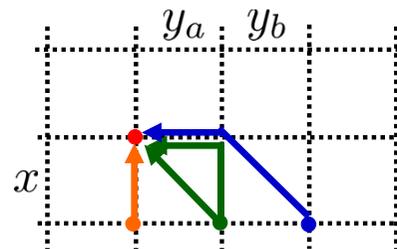
(1) Forward 確率



$$\alpha(\bullet) = \alpha(\bullet) \times p(\varepsilon|x) + \alpha(\bullet) \times p(y_b|x) + \alpha(\bullet) \times p(y_a y_b|x)$$

$$g_1(x; \mathbf{y})$$

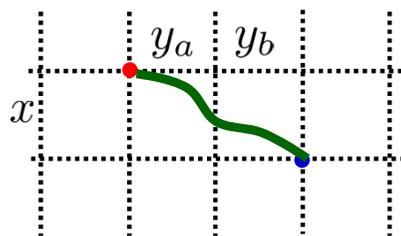
(2) Backward 確率



$$\beta(\bullet) = \beta(\bullet) \times p(\varepsilon|x) + \beta(\bullet) \times p(y_b|x) + \beta(\bullet) \times p(y_a y_b|x)$$

$$g_1(x; \mathbf{y})$$

(3) ビット尤度



$$p(\mathbf{y}|x_s = x) = \sum_{\bullet} \sum_{\bullet} \alpha(\bullet) p(\mathbf{y}|x) \beta(\bullet)$$

$$g_1(x; \mathbf{y})$$

通信路A :

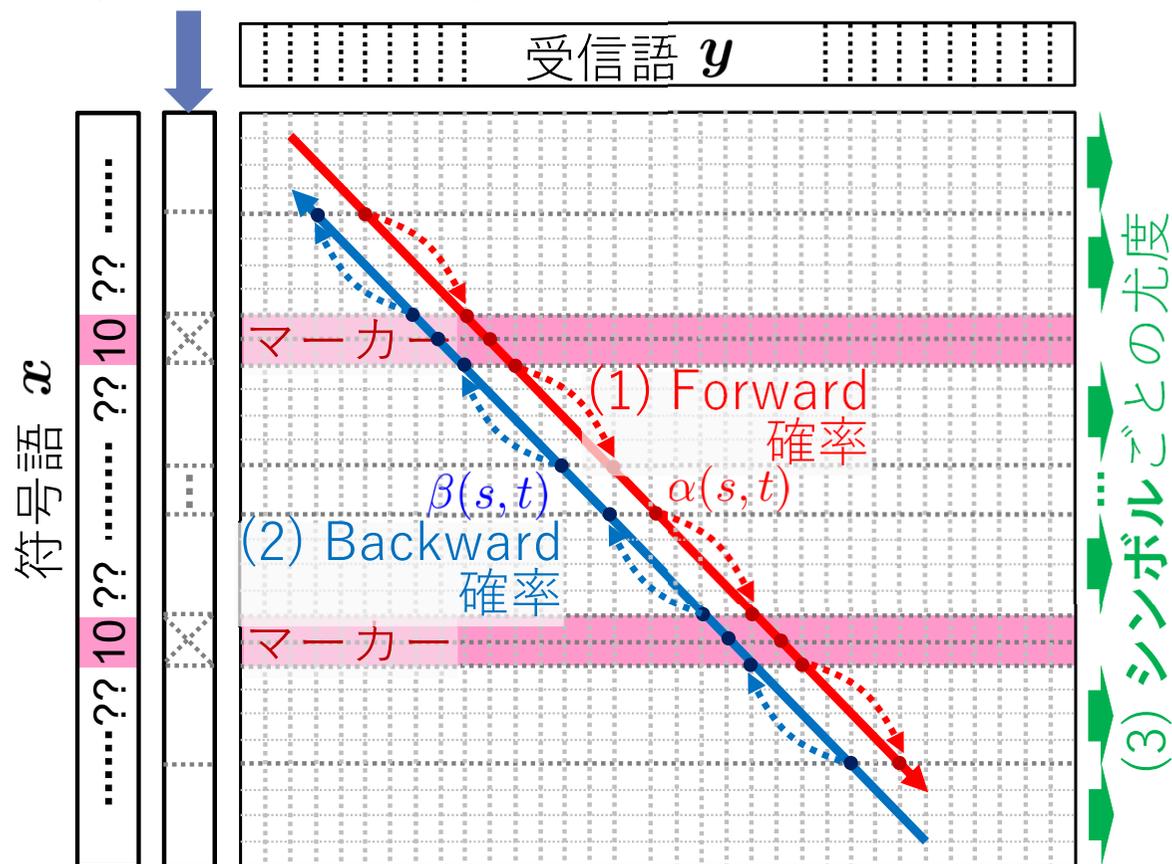
挿入: p_i 削除: p_d 反転: p_s
 挿入誤りの最大長: $l_I = 1$

$$g_1(x; \mathbf{y}) = p(\mathbf{y}|x) :$$

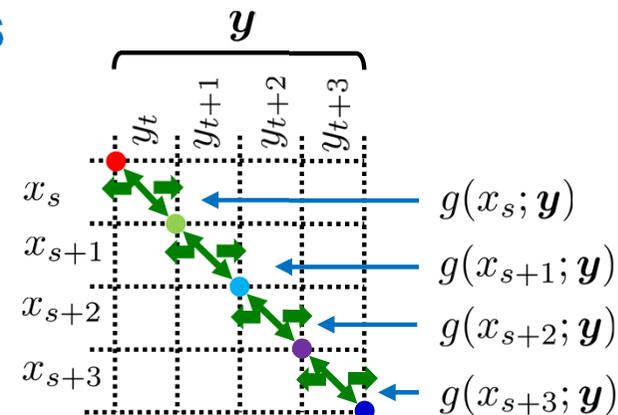
$x \in \mathbb{B}$ を送信したとき $\mathbf{y} = (y_0, \dots, y_{l-1})$ を受信する確率

Symbol-Level Synchronization (SLS)

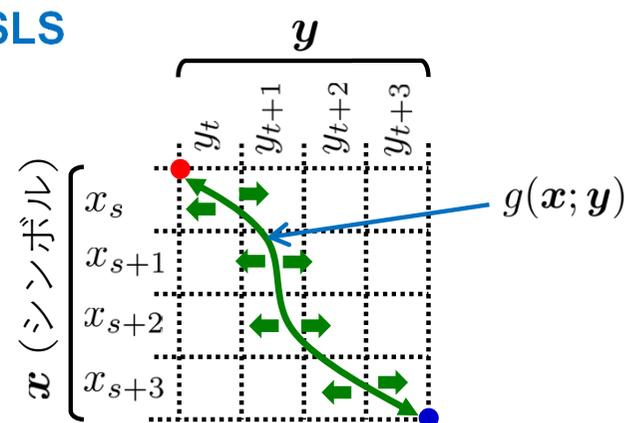
1 シンボル = b ビット
(フィードバック)



BLS



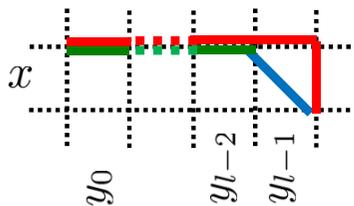
SLS



確率 $g(x; y)$ の再帰的な計算

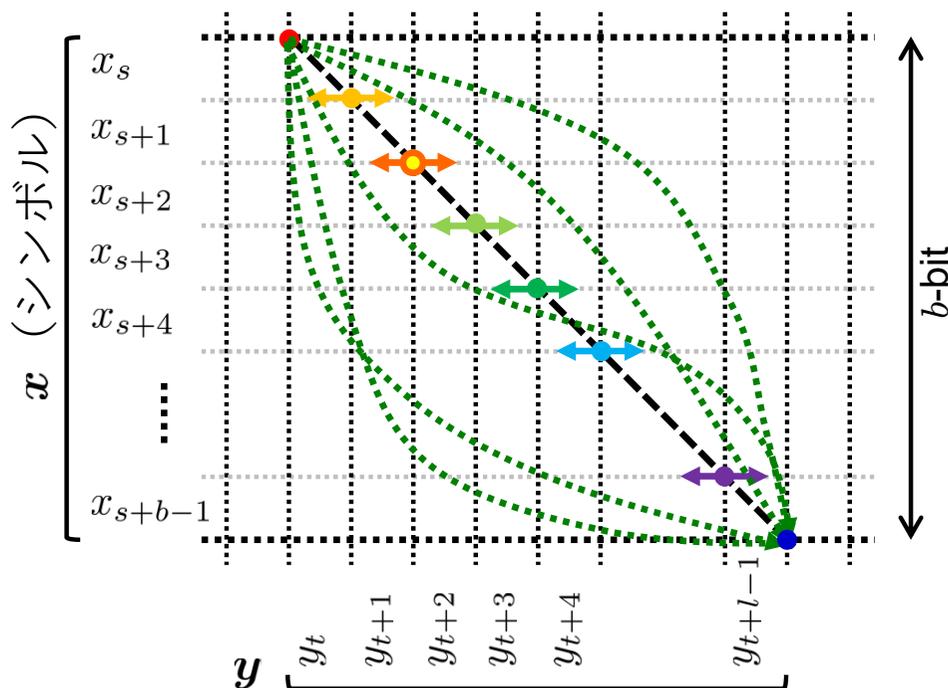
BLS: $x \in \mathbb{B}$ を送信したとき $\mathbf{y} = (y_0, \dots, y_{l-1}) \in \mathbb{B}^l$ を受信する確率

$$g_1(x; \mathbf{y}) = \underbrace{Q_{\text{ins}}(l-1)}_{\text{green}} p_t \underbrace{Q_{\text{sub}}(x; y_{l-1})}_{\text{blue}} + \underbrace{Q_{\text{ins}}(l)}_{\text{red}} p_d$$



SLS: $x \in \mathbb{B}^b$ を送信したとき $\mathbf{y} = (y_0, \dots, y_{l-1}) \in \mathbb{B}^l$ を受信する確率

$$g(\mathbf{x}; \mathbf{y}) = \sum_{\bullet} g_1(\cdot) \left(\sum_{\bullet} g_1(\cdot) g_1(\cdot) \right) \right) \right) \right) \right) \right) \right) \right) \right)$$



シンボルサイズ b に対し、
指数オーダで計算量が増加

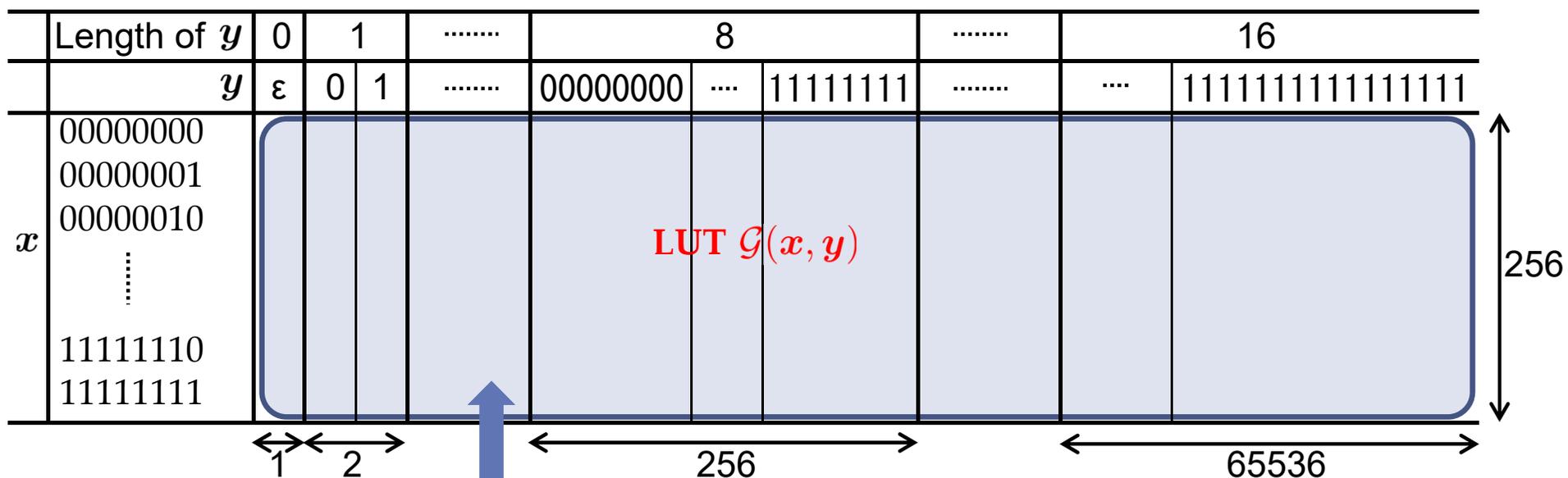
発表内容

- ◆ 研究の背景
- ◆ 挿入／削除／反転(IDS)誤り通信路
 - ◆ 通信路モデル
 - ◆ 従来のIDS誤り訂正符号の種類
 - ◆ マーカーを用いた符号化と復号
 - ◆ BLS/SLS
 - ◆ ルックアップテーブルによるSLSの高速化
- ◆ ドリフト通信路
 - ◆ 通信路モデル
 - ◆ マーカーを用いた符号化と復号
- ◆ まとめ

確率 $g(x; y)$ のルックアップテーブル

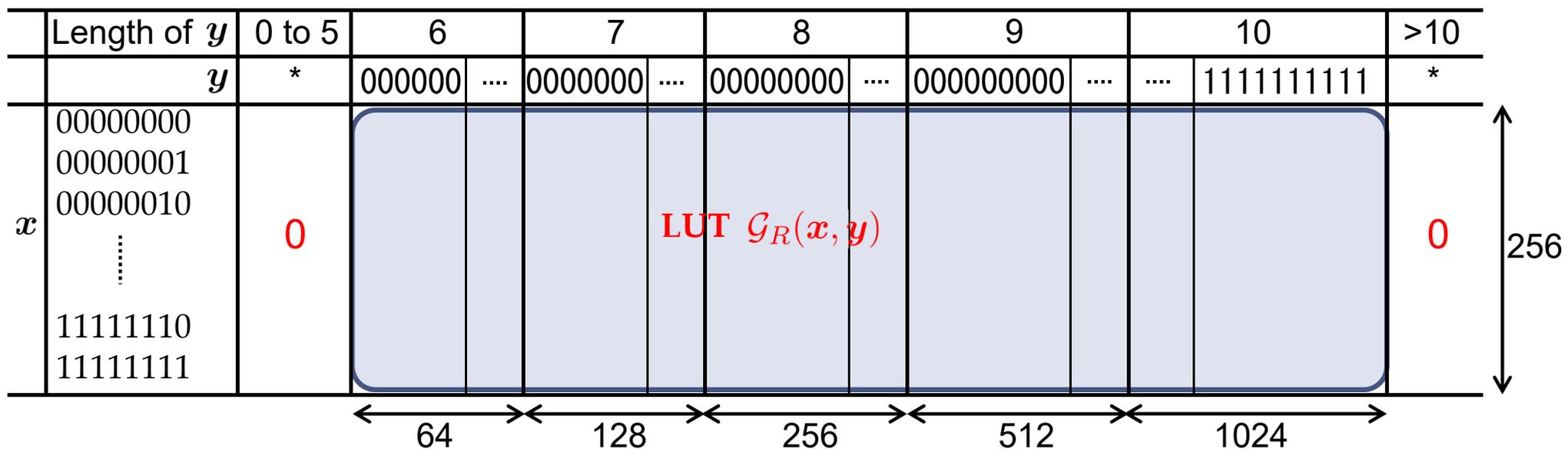
➡ 復号器のメモリに格納

シンボルサイズ: $b = 8$
 挿入誤りの最大長: $l_I = 1$



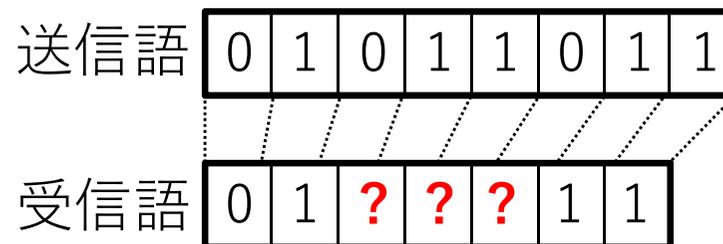
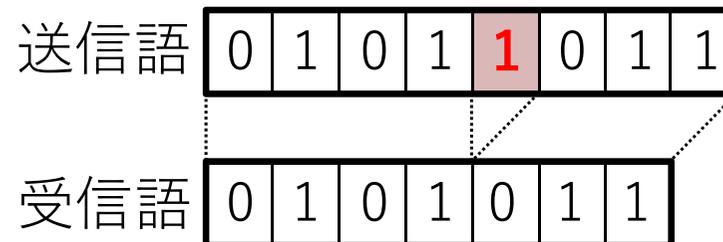
$$g(x; y) = \sum_{\bullet} g_1(\cdot) \left(\sum_{\circ} g_1(\cdot) \left(\sum_{\bullet} g_1(\cdot) g_1(\cdot) \right) \right) \right) \right) \right) \right) \right) \right)$$

ルックアップテーブルサイズの削減



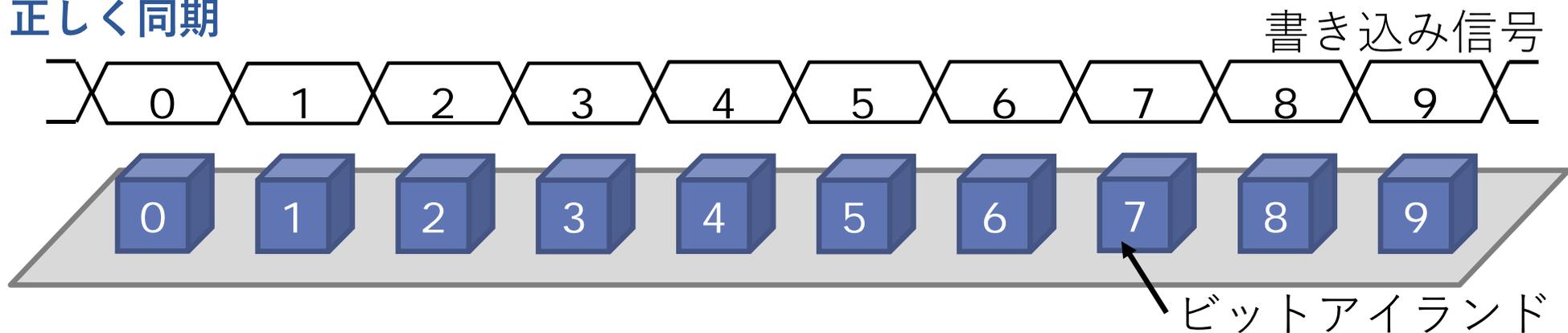
発表内容

- ◆ 研究の背景
- ◆ 挿入／削除／反転(IDS)誤り通信路
 - ◆ 通信路モデル
 - ◆ 従来のIDS誤り訂正符号の種類
 - ◆ マーカーを用いた符号化と復号
- ◆ ドリフト通信路
 - ◆ 通信路モデル
 - ◆ マーカーを用いた符号化と復号
- ◆ まとめ

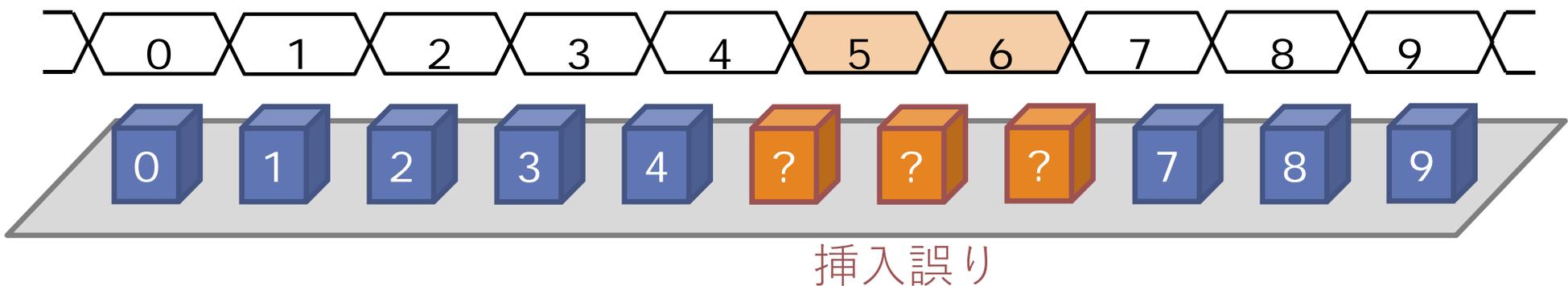


例：BPMで想定されるドリフト

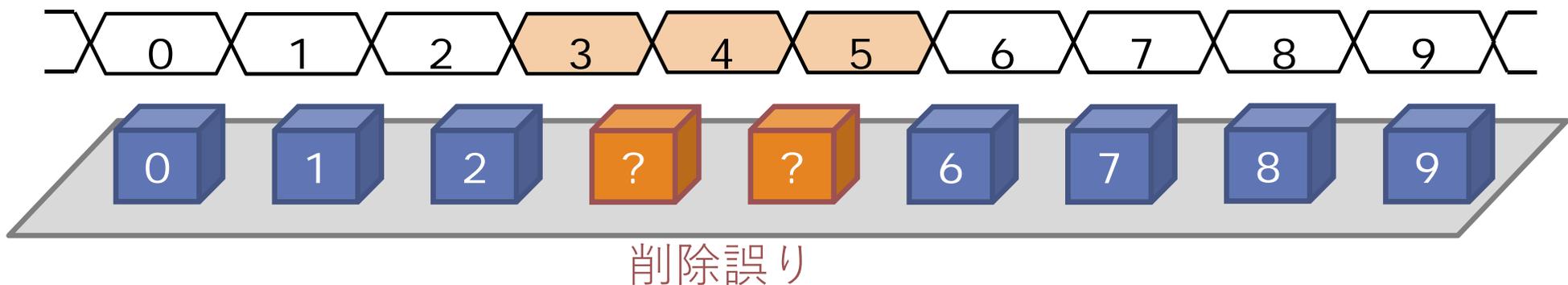
正しく同期



メディアが速い(または信号が遅い)

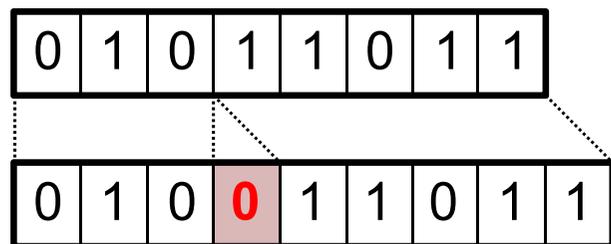


メディアが遅い(または信号が速い)

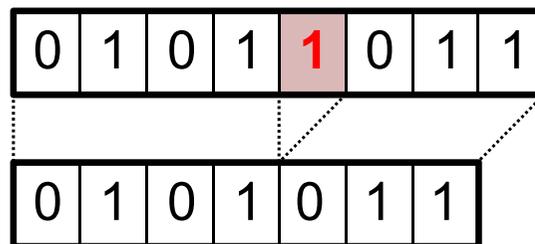


ドリフト通信路モデル

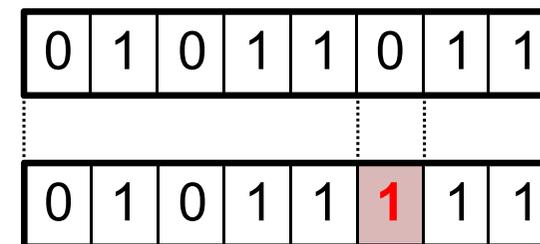
従来のIDS誤り通信路 ⇨ ビット単位での誤り



挿入誤り



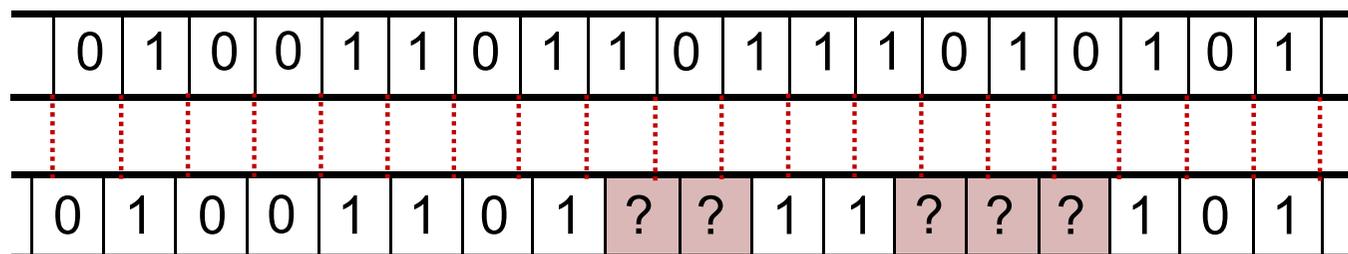
削除誤り



反転誤り

ドリフト通信路 ⇨ ビットの境界がずれる

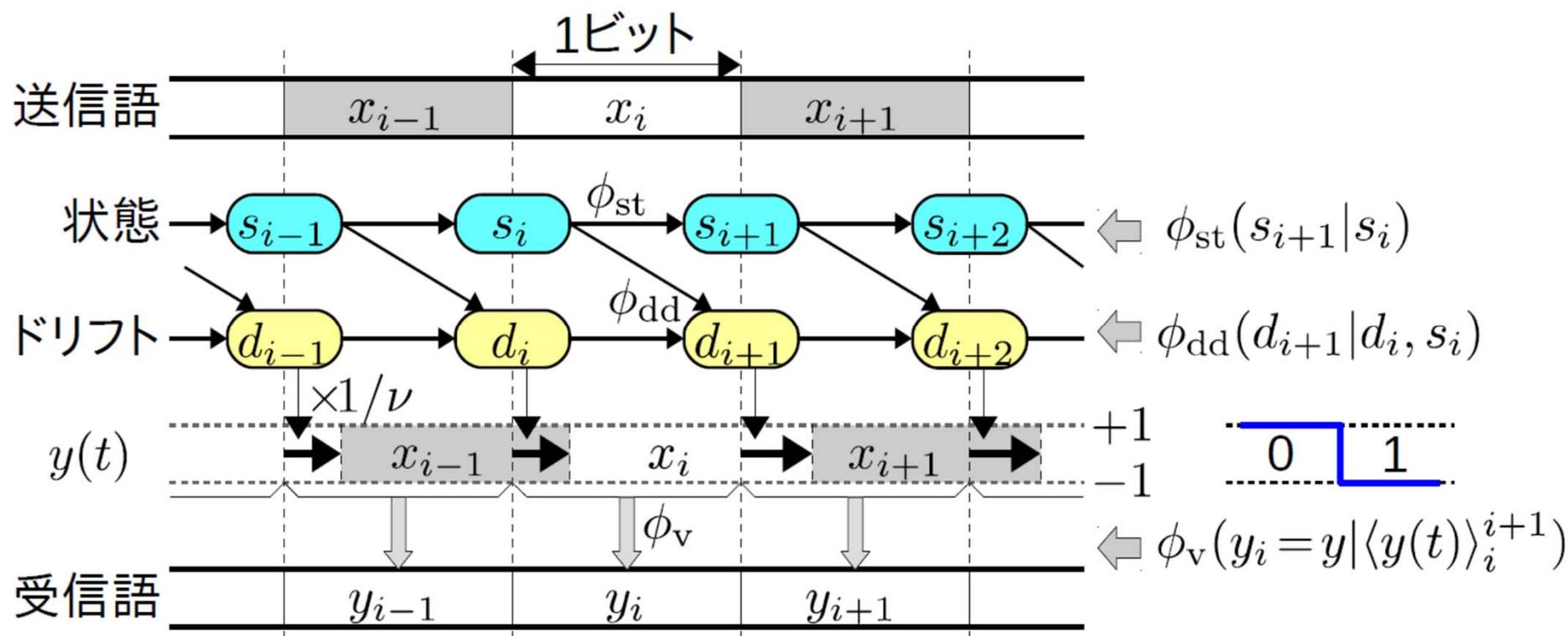
送信語



受信語

通信路モデル: 概要

送信語: $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ ($x_i, y_i \in \mathbb{B} = \{0, 1\}$)
 受信語: $\mathbf{y} = (y_0, y_1, \dots, y_{n-1})$



通信路モデル: 状態遷移

状態の有限集合: S

第 i ビット x_i を送信するときの状態: $s_i \in S$

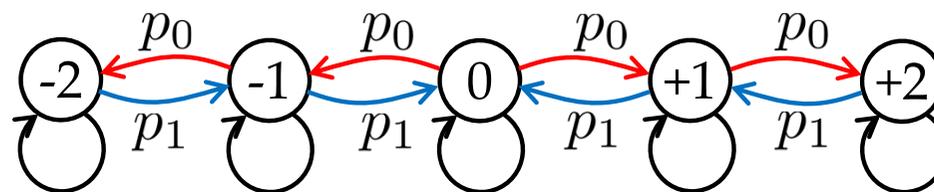
状態ベクトル: $\mathbf{s} = (s_0, s_1, \dots, s_{n-1}, s_n) \in S^{n+1}$
マルコフ過程

初期状態 s_0 の PMF: $\phi_{\text{is}}(s_0)$
状態遷移確率: $\phi_{\text{st}}(s'|s)$ } channel parameter

例: $S = \{-2, -1, 0, 1, 2\}$

$$\phi_{\text{is}}(s_0) = \begin{cases} 1 & (s_0 = 0) \\ 0 & (s_0 \in \{-2, -1, 1, 2\}) \end{cases} \quad (\text{初期状態} = 0)$$

$\phi_{\text{st}}(s'|s)$:

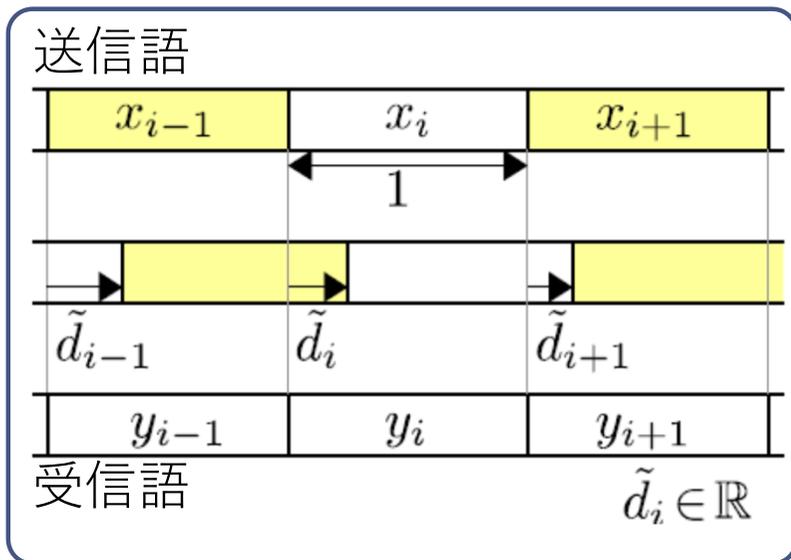


(状態: ドリフト値の勾配)

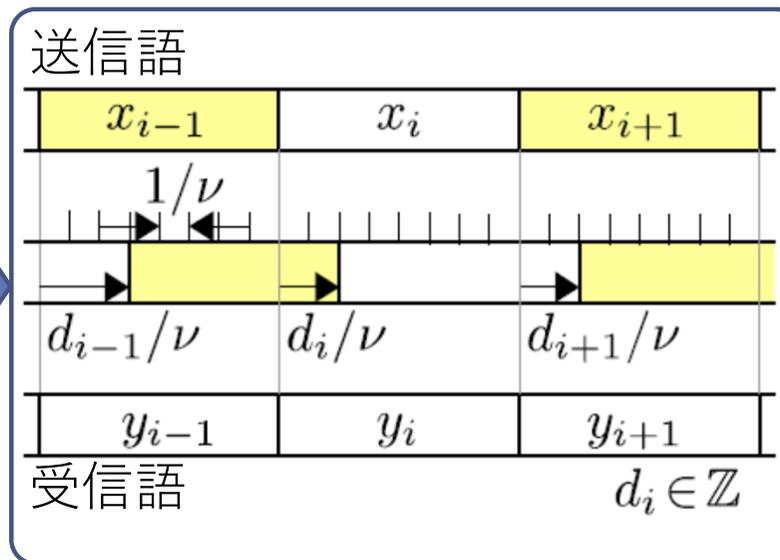
通信路モデル: ドリフト

第 i ビット x_i のドリフト: $\tilde{d}_i \in \mathbb{R} \longrightarrow d_i/\nu \quad (d_i \in \mathbb{Z})$
 └─ 量子化した値で近似

ドリフト値 (実数)



量子化したドリフト値

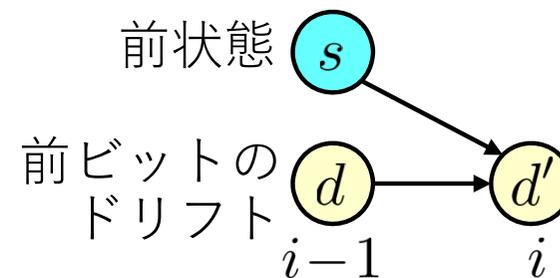


1-bit = ν セグメント

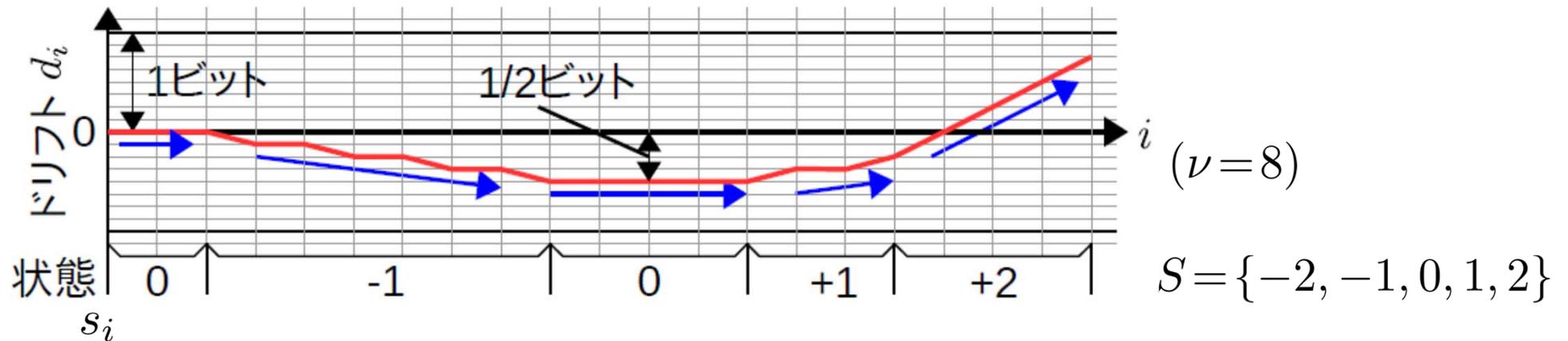
ドリフトベクトル: $\mathbf{d} = (d_0, d_1, \dots, d_{n-1}, d_n)$

ドリフト初期値のPMF: $\phi_{\text{id}}(d_0)$

ドリフト値のPMF: $\phi_{\text{dd}}(d' | d, s)$



通信路モデル: ドリフトモデルの例



ドリフト初期値のPMF: $\phi_{id}(d_0) = \frac{c}{\sqrt{2\pi\sigma_d^2}} \exp\left(\frac{-d_0^2}{2\sigma_d^2}\right)$

平均: 0 分散: σ_d^2

ドリフト値のPMF: $\phi_{dd}(d'|d, s) = \frac{c}{\sqrt{2\pi\sigma_\delta^2}} \exp\left(\frac{-(d' - (d + \gamma s))^2}{2\sigma_\delta^2}\right)$

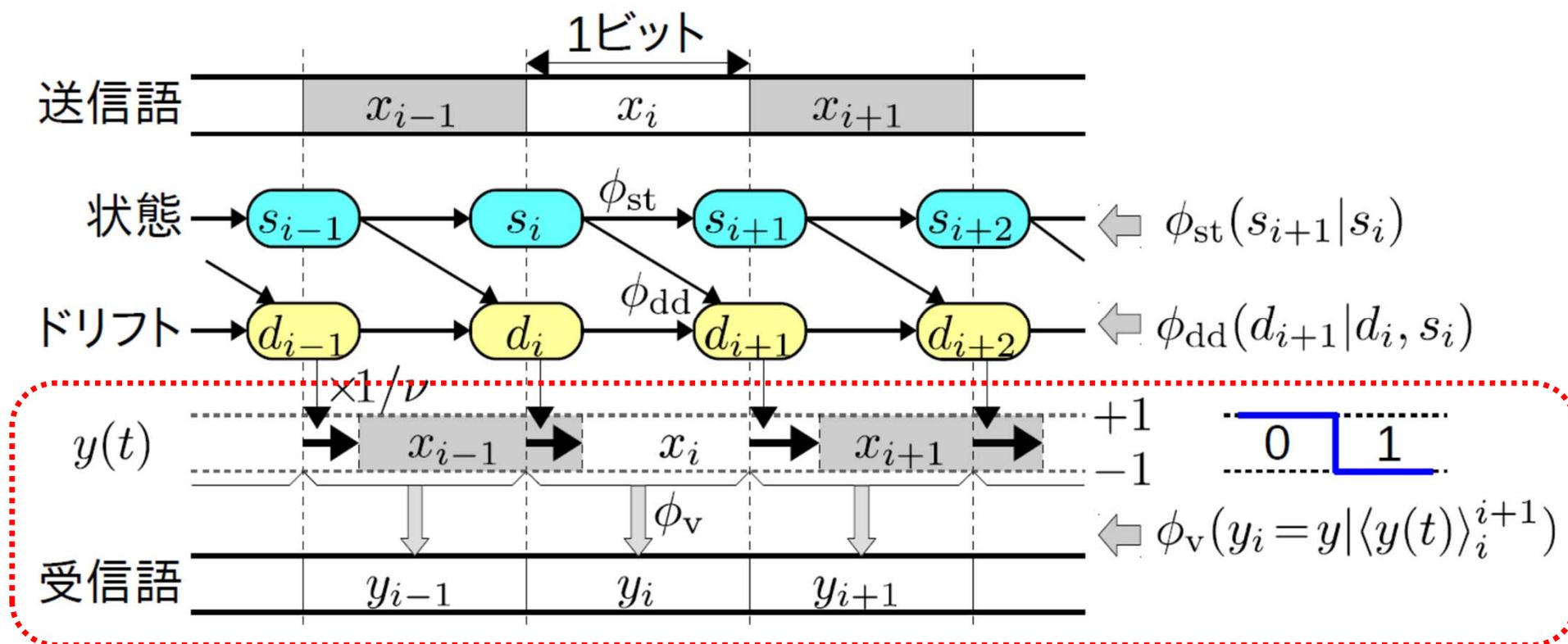
平均: $d + \gamma s$ 分散: σ_δ^2

↑ ↑

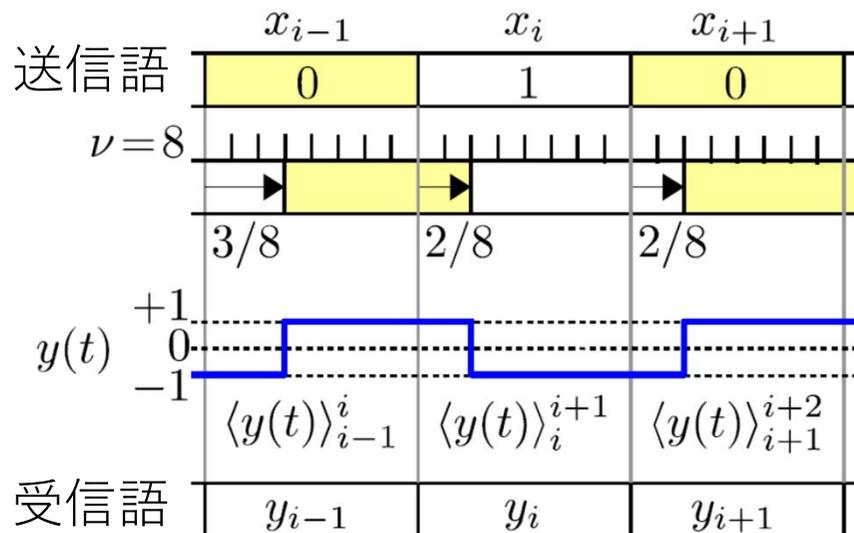
前ビットのドリフト 前状態

通信路モデル: 概要

送信語: $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$
 受信語: $\mathbf{y} = (y_0, y_1, \dots, y_{n-1})$ ($x_i, y_i \in \mathbb{B} = \{0, 1\}$)



通信路モデル: 受信ビットの値



$$(\dots, x_{i-1}, x_i, x_{i+1}, \dots) = (\dots, 0, 1, 0, \dots)$$

$$(\dots, d_{i-1}, d_i, d_{i+1}, \dots) = (\dots, 3, 2, 2, \dots)$$

$$y(t) = \begin{cases} -1^{x_i} & (i + \frac{d_i}{\nu} \leq t < i + 1 + \frac{d_{i+1}}{\nu}) \\ 0 & (\text{otherwise}) \end{cases}$$

$$\phi_{\nu}(y_i | \langle y(t) \rangle_i^{i+1})$$

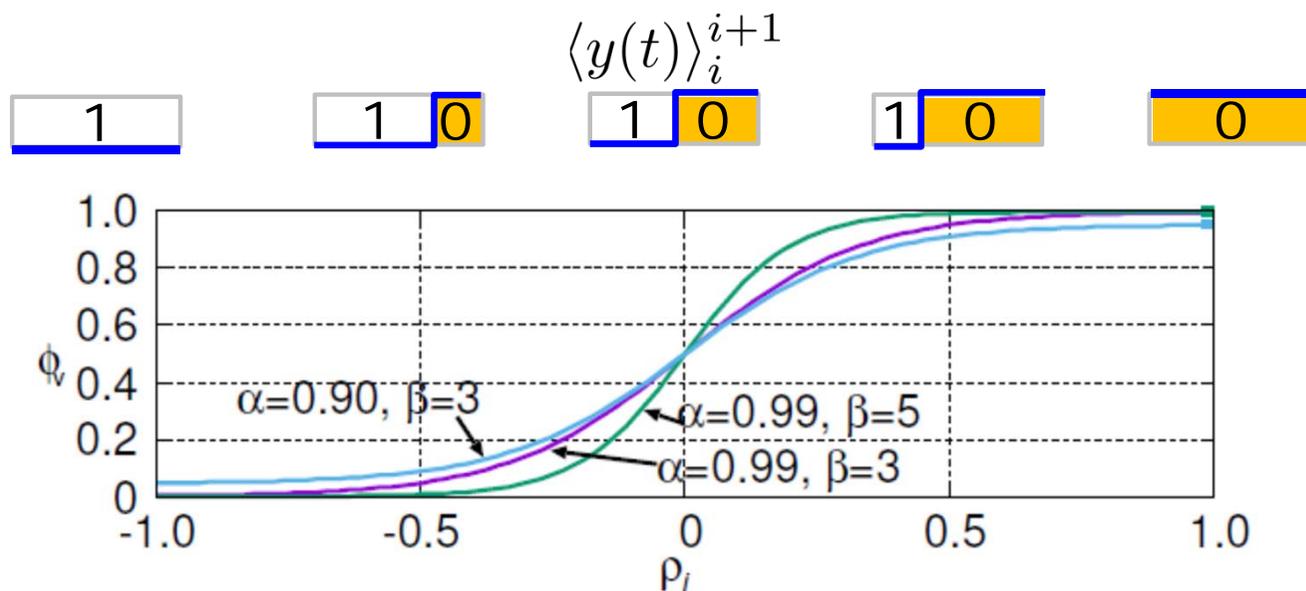
例:

$$\phi_{\nu}(y_i = 0 | \langle y(t) \rangle_i^{i+1})$$

$$= \frac{1}{2}(\alpha \tanh(\beta \rho) + 1)$$

$$\rho = \int_i^{i+1} y(t) dt$$

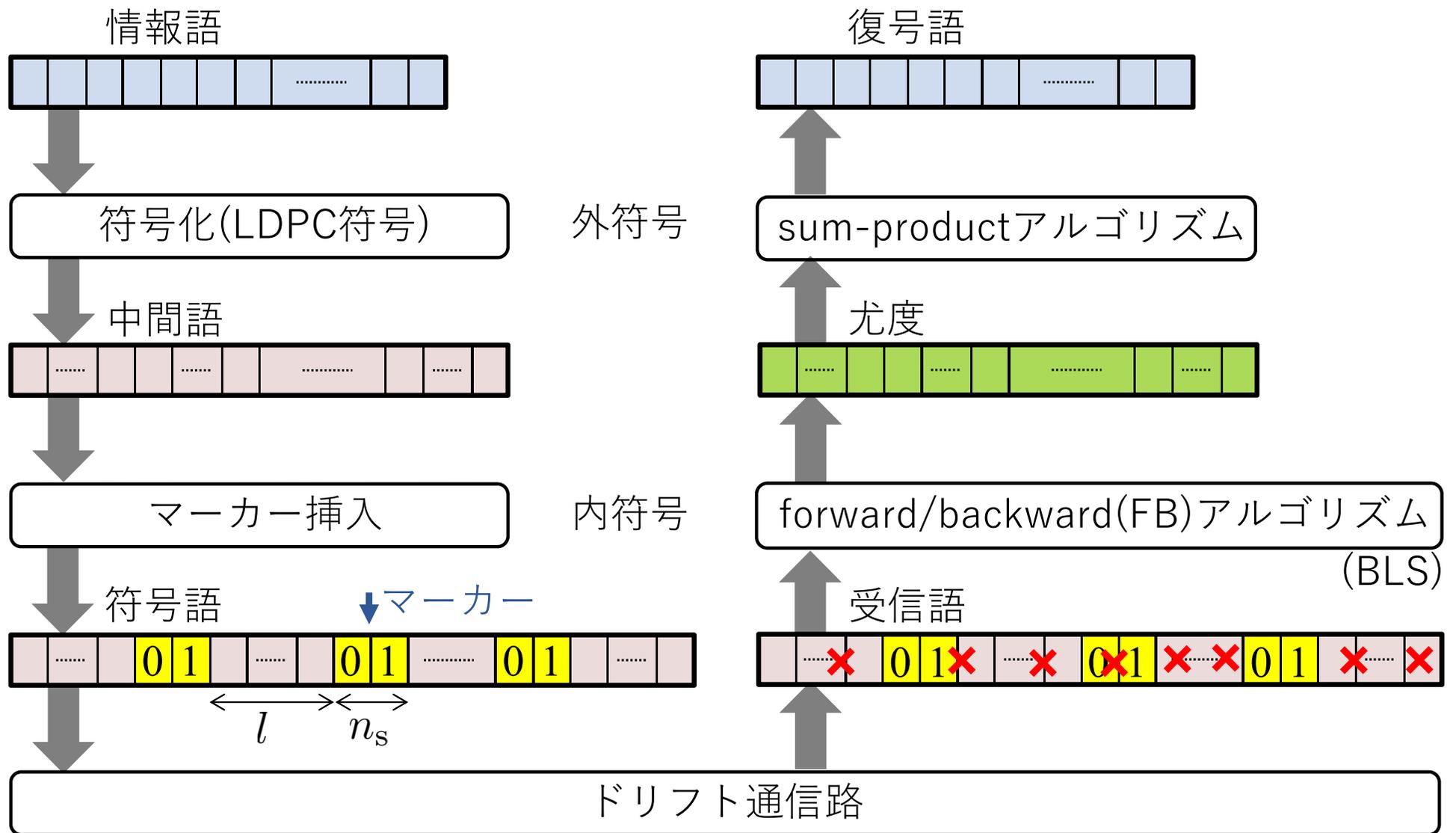
α, β : パラメータ



発表内容

- ◆ 研究の背景
- ◆ 挿入／削除／反転(IDS)誤り通信路
 - ◆ 通信路モデル
 - ◆ 従来のIDS誤り訂正符号の種類
 - ◆ マーカーを用いた符号化と復号
- ◆ ドリフト通信路
 - ◆ 通信路モデル
 - ◆ マーカーを用いた符号化と復号
- ◆ まとめ

符号化と復号



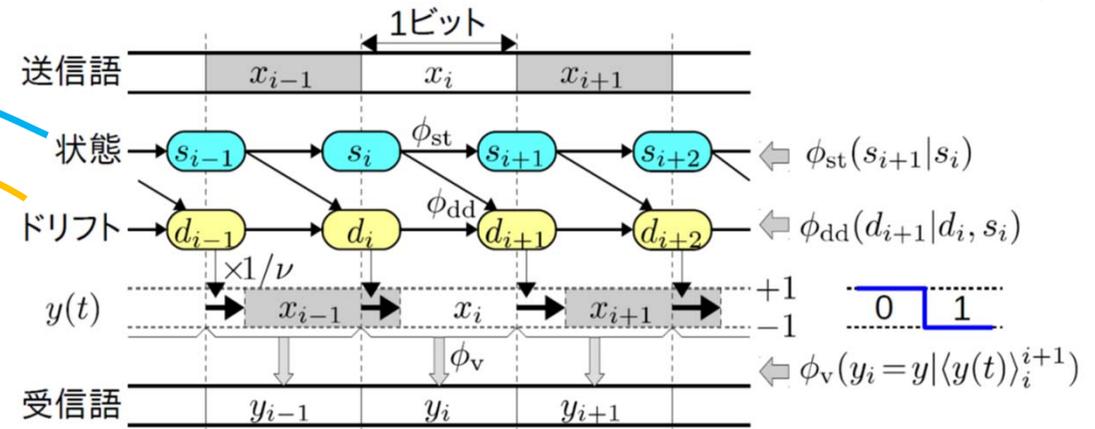
FBアルゴリズムの概要

通信路の状態集合: S

ドリフト値の集合:

$$D = \{-d^w, \dots, -1, 0, 1, \dots, d^w\}$$

d^w : ドリフト最大値

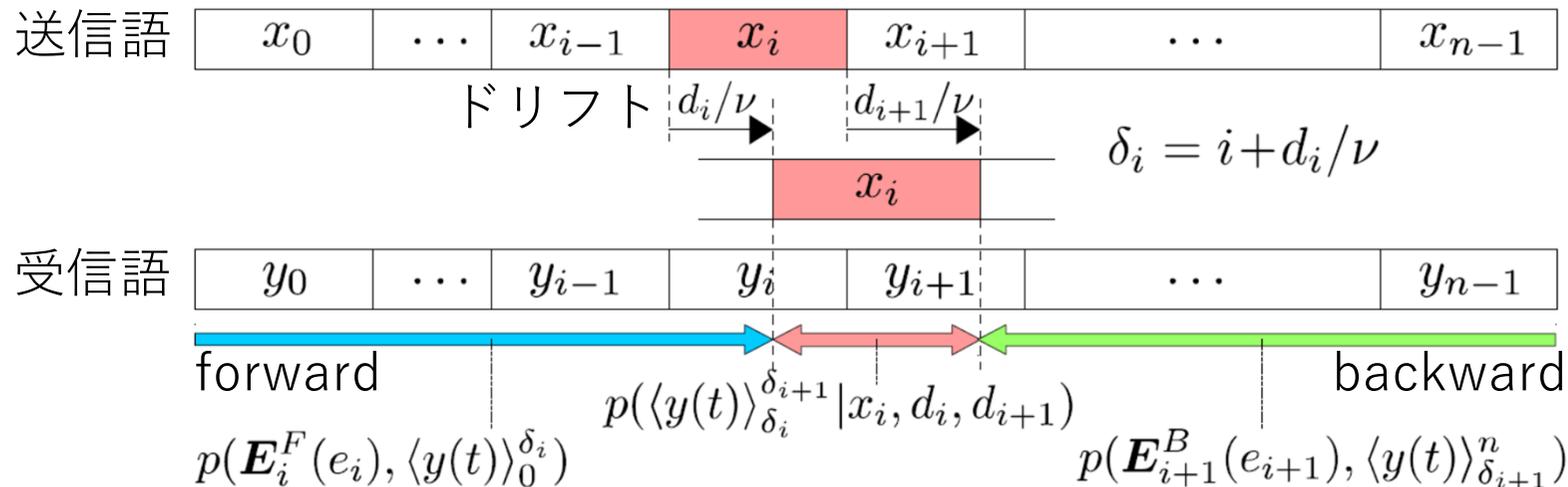


state-drift vector (SDV): $e = (e_0, e_1, \dots, e_n) \in E^{n+1}$ ($e_i \in E = S \times D$)

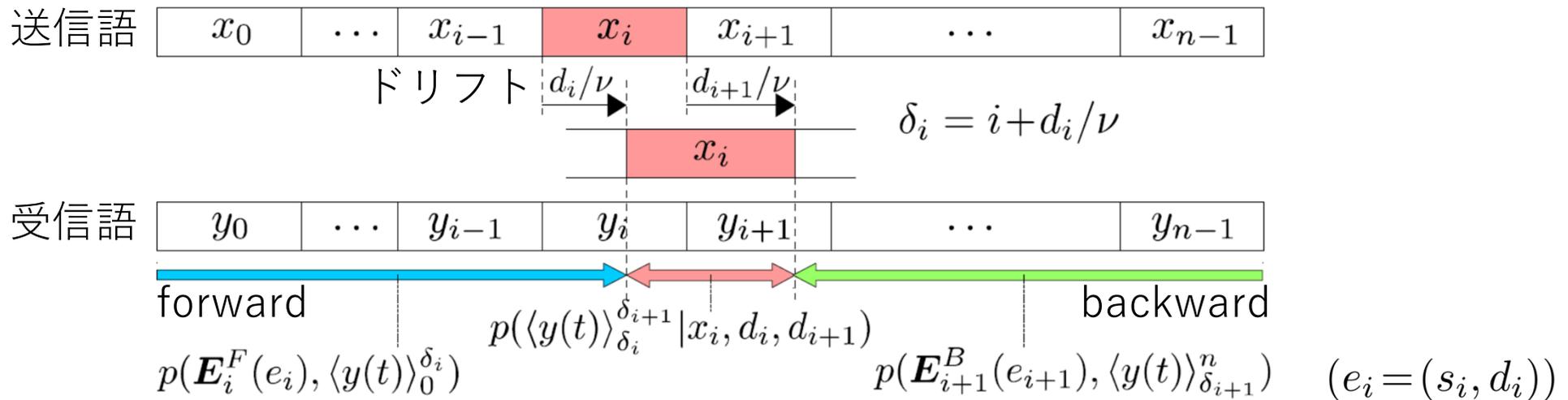
set of sub-sequences of SDV: $\mathbf{E}_i^F(e) = \{(e_0, \dots, e_i) \in E^{i+1} | e_i = e\}$

$$\mathbf{E}_i^B(e) = \{(e_i, \dots, e_n) \in E^{n-i+1} | e_i = e\}$$

FBアルゴリズム:



FBアルゴリズム



forward:

$$\Pr(\mathbf{E}_i^F(e_i), \langle y(t) \rangle_0^{\delta_i}) \propto \sum_{x \in \mathbb{B}} \Pr(x_{i-1} = x) \sum_{e_{i-1} \in E} \Pr(\mathbf{E}_{i-1}^F(e_{i-1}), \langle y(t) \rangle_0^{\delta_{i-1}}) \Pr(e_i | e_{i-1}) \Pr(\langle y(t) \rangle_{\delta_{i-1}}^{\delta_i} | x_{i-1} = x, d_{i-1}, d_i)$$

backward:

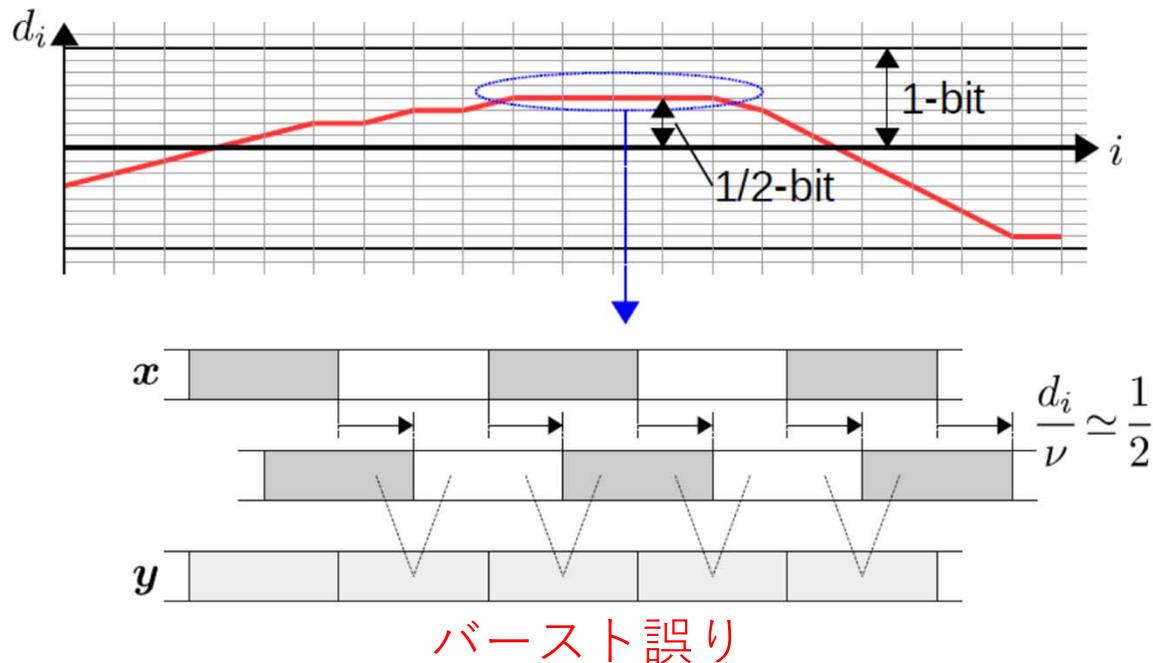
$$\Pr(\mathbf{E}_i^B(e_i), \langle y(t) \rangle_{\delta_i}^n) \propto \sum_{x \in \mathbb{B}} \Pr(x_i = x) \sum_{e_{i+1} \in E} \Pr(\mathbf{E}_{i+1}^B(e_{i+1}), \langle y(t) \rangle_{\delta_{i+1}}^n) \Pr(e_{i+1} | e_i) \Pr(\langle y(t) \rangle_{\delta_i}^{\delta_{i+1}} | x_i = x, d_i, d_{i+1})$$

尤度:

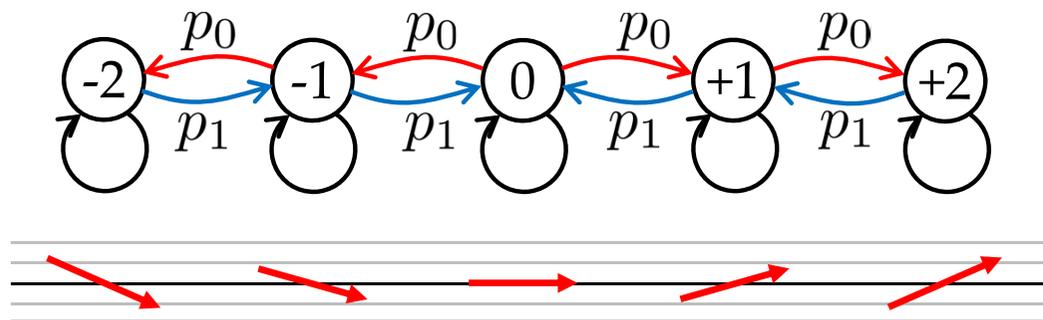
$$\Pr(y(t) | x_i = x) \propto \sum_{e_i \in E} \sum_{e_{i+1} \in E} \Pr(\mathbf{E}_i^F(e_i), \langle y(t) \rangle_0^{\delta_i}) \Pr(\langle y(t) \rangle_{\delta_i}^{\delta_{i+1}} | x, d_i, d_{i+1}) \Pr(e_{i+1} | e_i) \Pr(\mathbf{E}_{i+1}^B(e_{i+1}), \langle y(t) \rangle_{\delta_{i+1}}^n)$$

1/2ビットドリフトによるバースト誤り

ドリフト値が $\frac{d_i}{\nu} = \frac{1}{2}$ の周辺 \rightarrow 受信ビット：高い誤り率

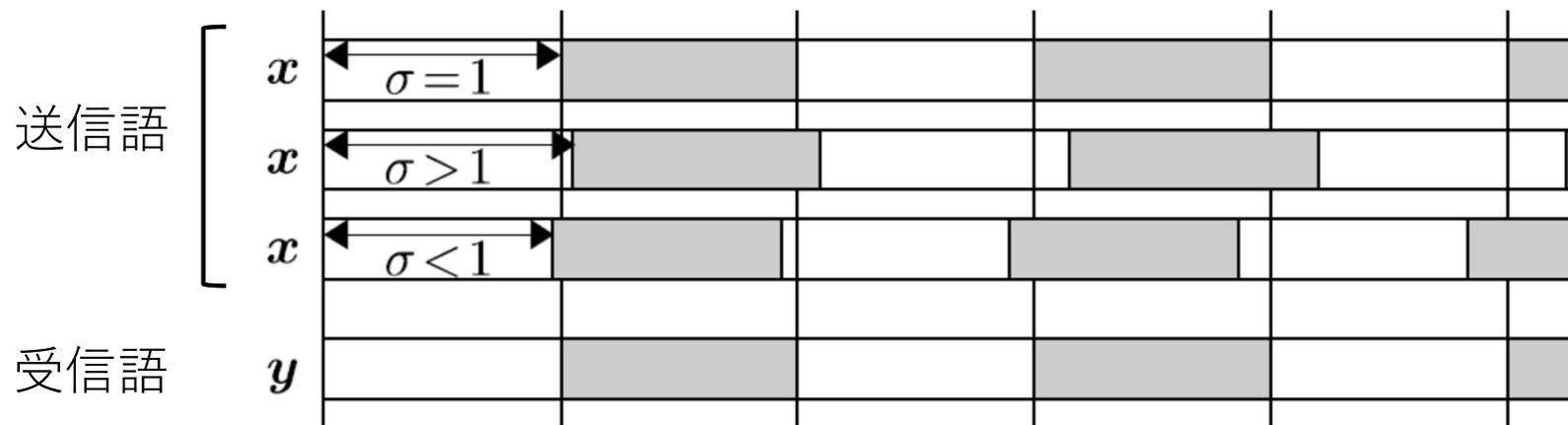


$p(d_i = d_{i+1})$ が大きいとき \rightarrow 長いバースト誤り



ビット幅のスケーリング

送信語のビットの「幅」を $\sigma \simeq 1$ 倍にスケール



$\sigma > 1$ \Rightarrow 挿入誤りの発生

$\sigma < 1$ \Rightarrow 削除誤りの発生

(同期を意図的にずらして長いバースト誤りを避ける)

まとめ

- ◆ 研究の背景：同期誤り
 - ◆ BPM, racetrack memory, DNA storage
- ◆ 挿入／削除／反転(IDS)誤り通信路
 - ◆ 通信路モデル
 - ◆ 従来のIDS誤り訂正符号の種類
 - ◆ マーカーを用いた符号化と復号
- ◆ ドリフト通信路
 - ◆ 通信路モデル
 - ◆ マーカーを用いた符号化と復号

