

データの誤り検出・訂正と データ量を減らすしくみ

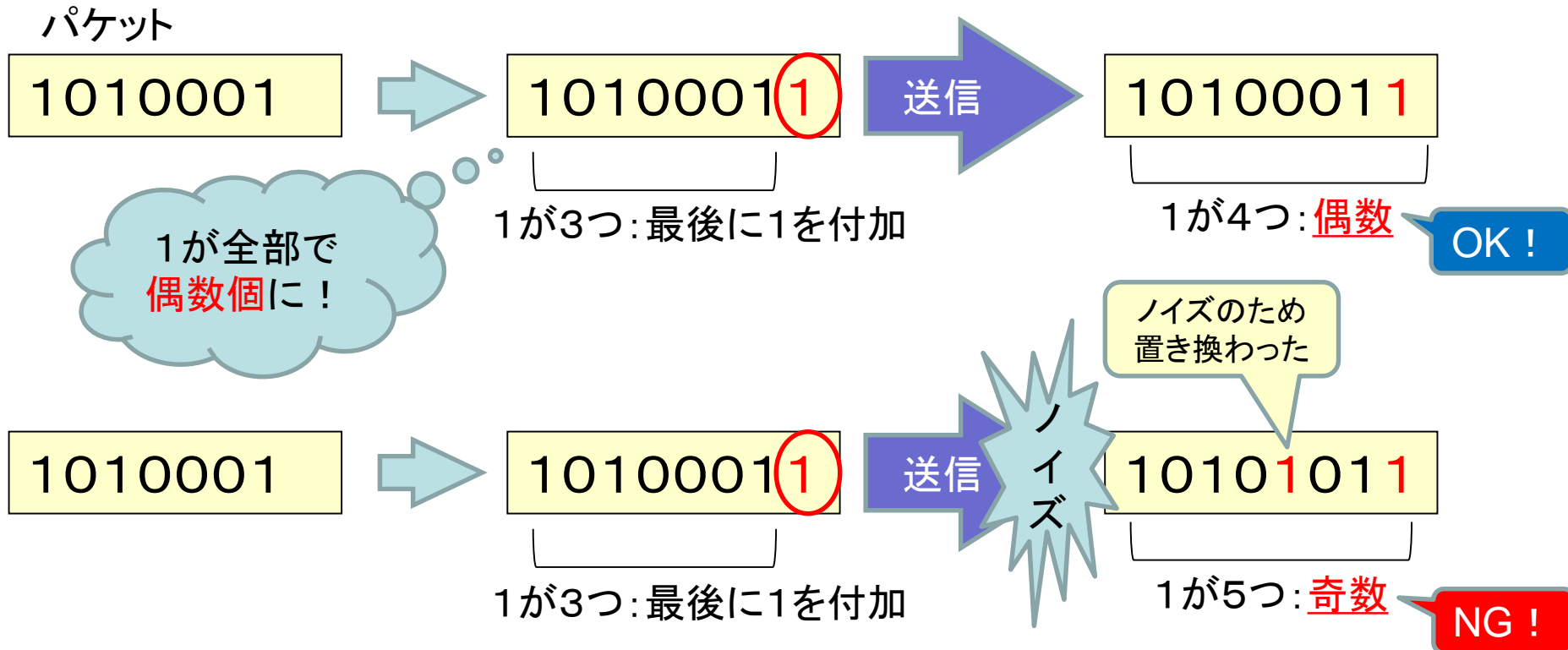
情報の科学 第29回授業

05コンピュータによる情報の処理

対応ファイル: 21exp29.xls

データの誤り検出、訂正

1の数が奇数個の場合は「1」を付加し、偶数個の場合は「0」を付加して、1の数が全体で偶数個になるようにする。(垂直パリティチェック)



※偶数個にすることを「偶数パリティチェック」という。

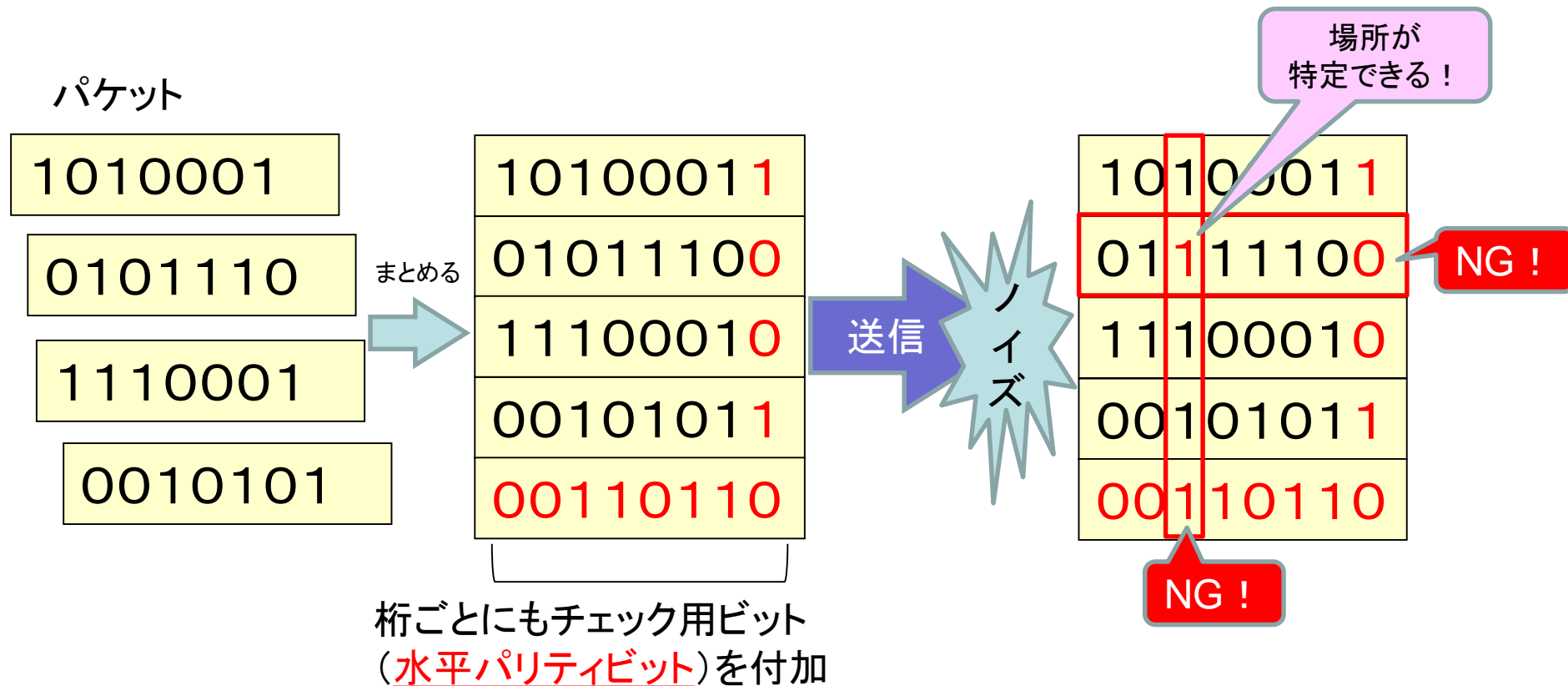
(奇数個にする「奇数パリティチェック」もある)

※垂直パリティチェックでは、「誤り」があることはわかるが、どこが誤っているかまではわからない!

破棄して再送信!

データの誤り検出、訂正

複数のビット列をまとめ、データ列の桁ごとに、さらにチェック用ビットを追加することによって、誤りの箇所を特定し、訂正することができる。(水平パリティチェック)



※複数の誤りがある場合、それが違うビット列や桁であれば、ある程度絞り込めるが、同じビット列や同じ桁に複数の誤りがある場合、特定することはできない。

演習

- ワークシート練習1～3

データの圧縮

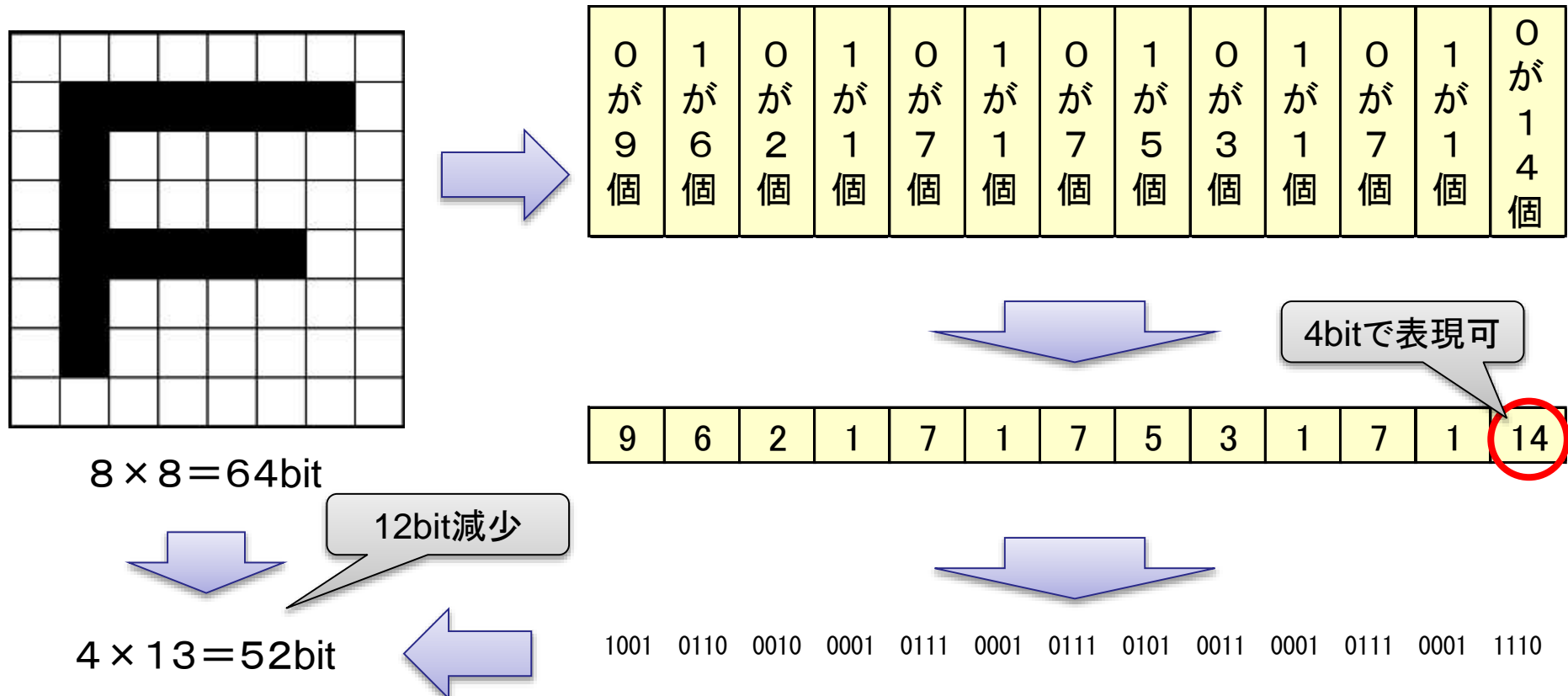
- 可逆圧縮
 - 完全に元の情報を復元する圧縮方法
 - GIF、PNGなどの圧縮、ファイルの圧縮
- 非可逆圧縮
 - 完全には元の情報に戻らない圧縮方法
 - JPEG、MP3などの圧縮

※一般に、非可逆圧縮の方が可逆圧縮よりも圧縮率が高い

可逆圧縮のしくみ(1)

- ランレングス圧縮

- 同じデータの繰り返しパターンに注目した方式



※圧縮するものによって圧縮率は変わり、場合によっては、むしろ増えてしまうこともある！

演習

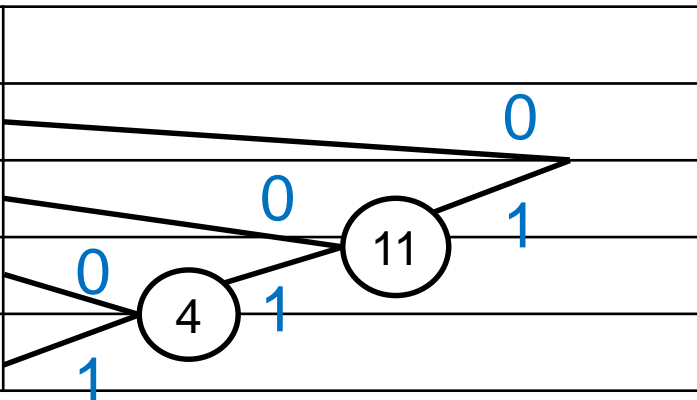
- ワークシート練習4

可逆圧縮のしくみ(2)

- ハフマン圧縮

- 出現頻度が高いデータを短いビット列に割り当てる。

例) AABAAACAABADABBBC

文字	頻度		符号
A	9		0
B	7		10
C	3		110
D	1		111

5. 道順をたどっていった「0」と「1」の符号を文字に割り当てる
4. 新たにできた「節」の値を1つの文字の頻度に見立て、1, 2を繰り返す
3. 2つの頻度の和(この場合は1+3)を求め新たな「節」を作る
2. 一番頻度の少ない2つを取り出し、0と1に振り分ける
(どちらを1とするかは任意だが、下を1とすると分かりやすい)
1. 頻度(確率)が高い順に並び替える

練習

- ワークシート練習5

ファイルの圧縮

- 圧縮ソフトウェアを用い、ファイル自体を圧縮
 - ZIP方式 …… windows標準対応(右クリック)
 - LZH方式 …… パソコン通信時代によく利用
 - RAR方式 …… データ破損にある程度まで対応

実習

- 圧縮用の2つのファイルをそれぞれ圧縮し、圧縮前後でファイルサイズを比較してみよう