Beyond Single-Deletion Correcting Codes:

Substitutions and Transpositions

Ryan Gabrys, Venkatesan Guruswami, João Ribeiro, **Ke Wu**

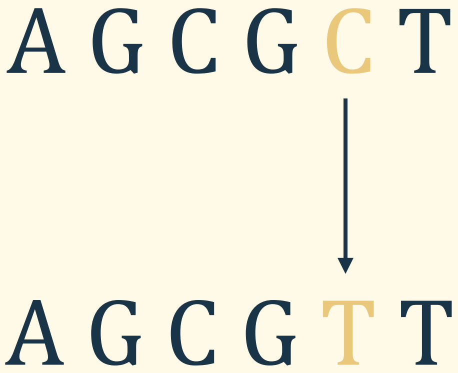
Three types of common errors

Substitution

Deletion(insertion)

Transposition

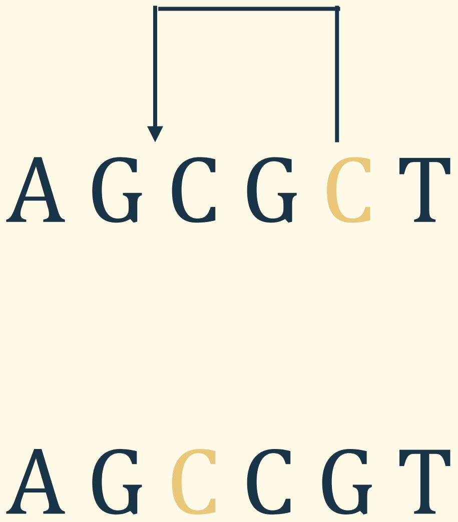
Substitution



Deletion(insertion)

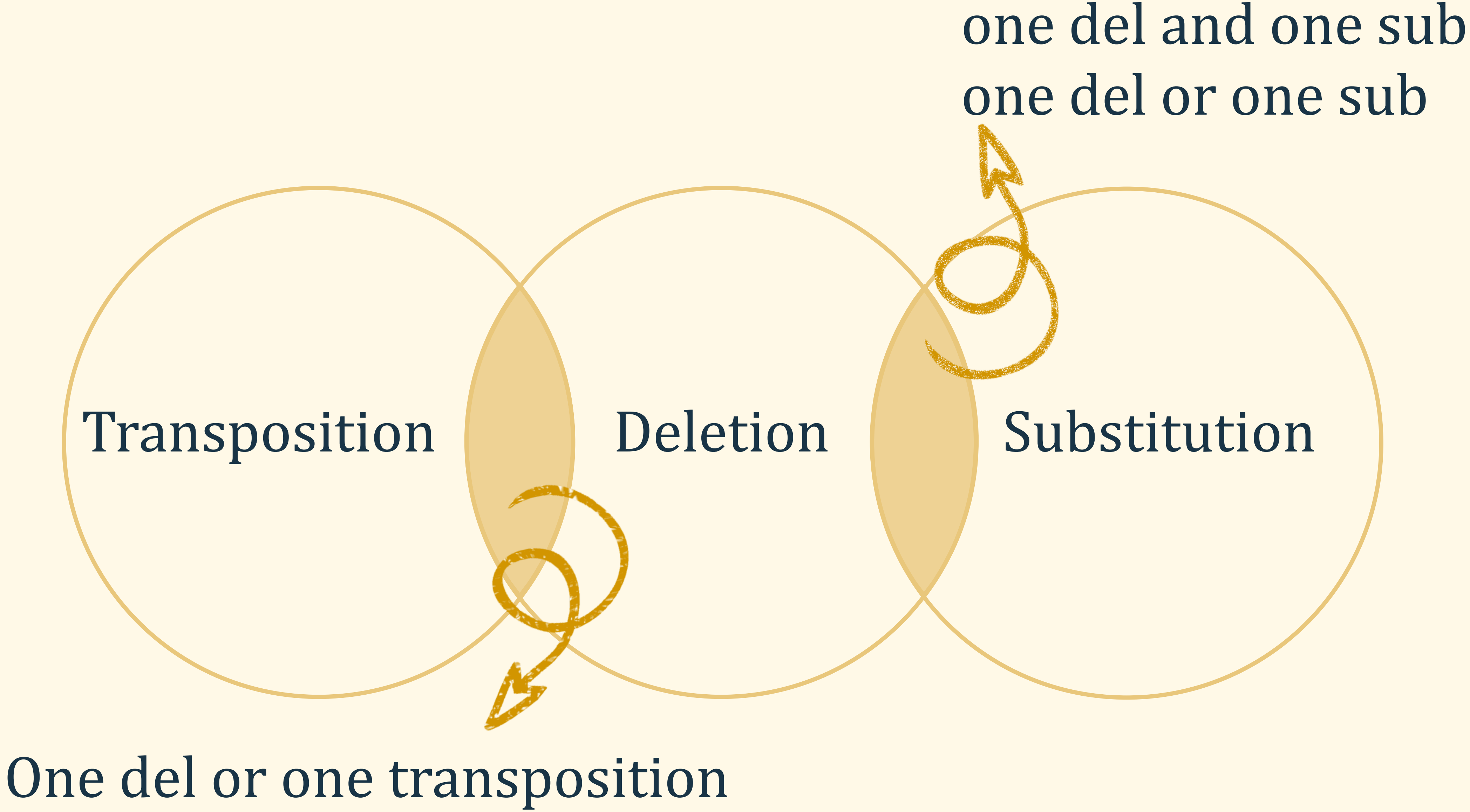
A G C G C T

A G C G T



Transposition

How about the interplay?

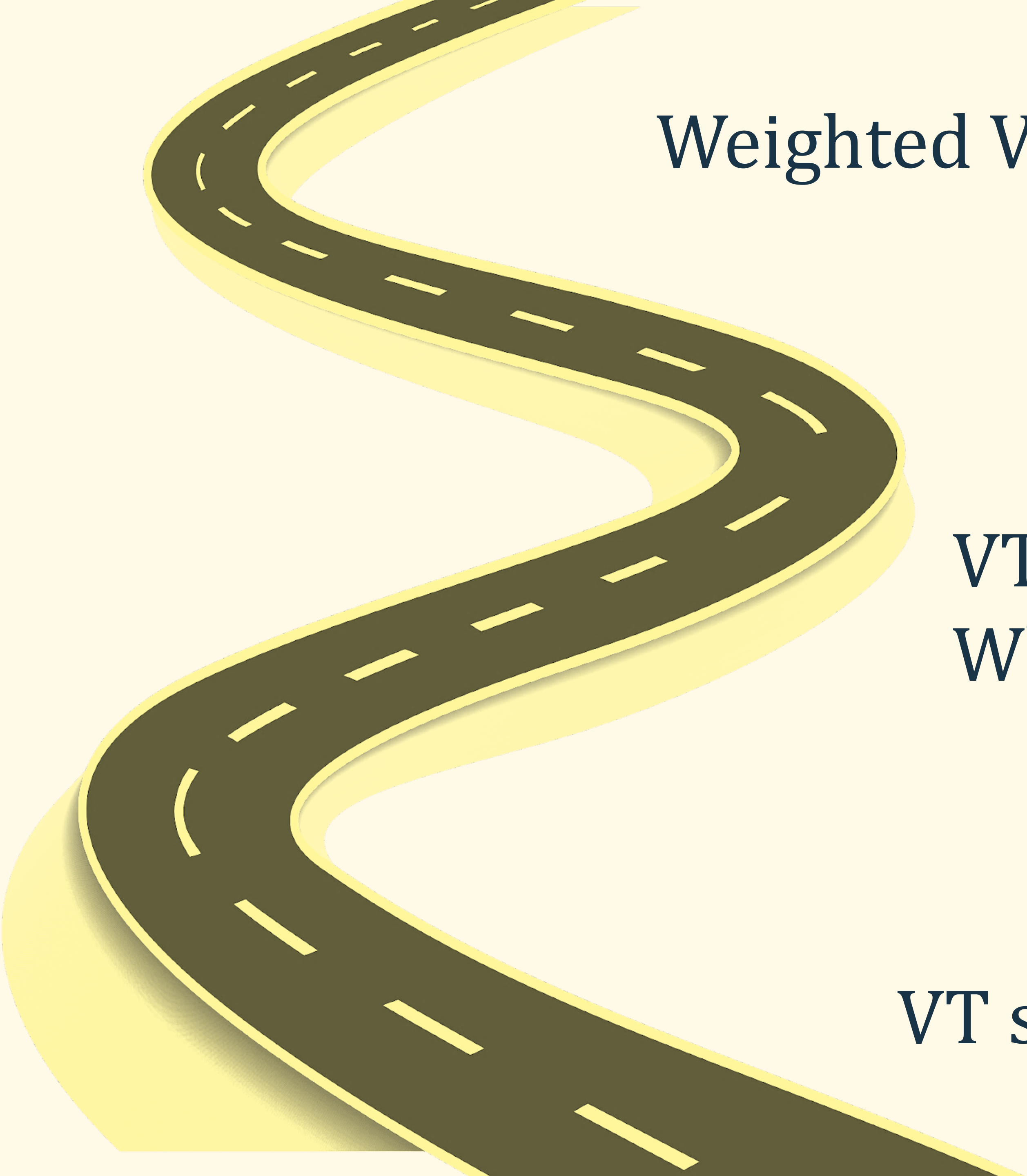


Our Results

Alphabet	Error type	Redundancy
q	One del or one sub (edit error)	$\log n + O_q(\log\log n)$
2	One del or one adjacent trans	$\log n + O(\log\log n)$
1	One del AND one sub	$4\log n + O(\log\log n)$ List of size 2

Our Results

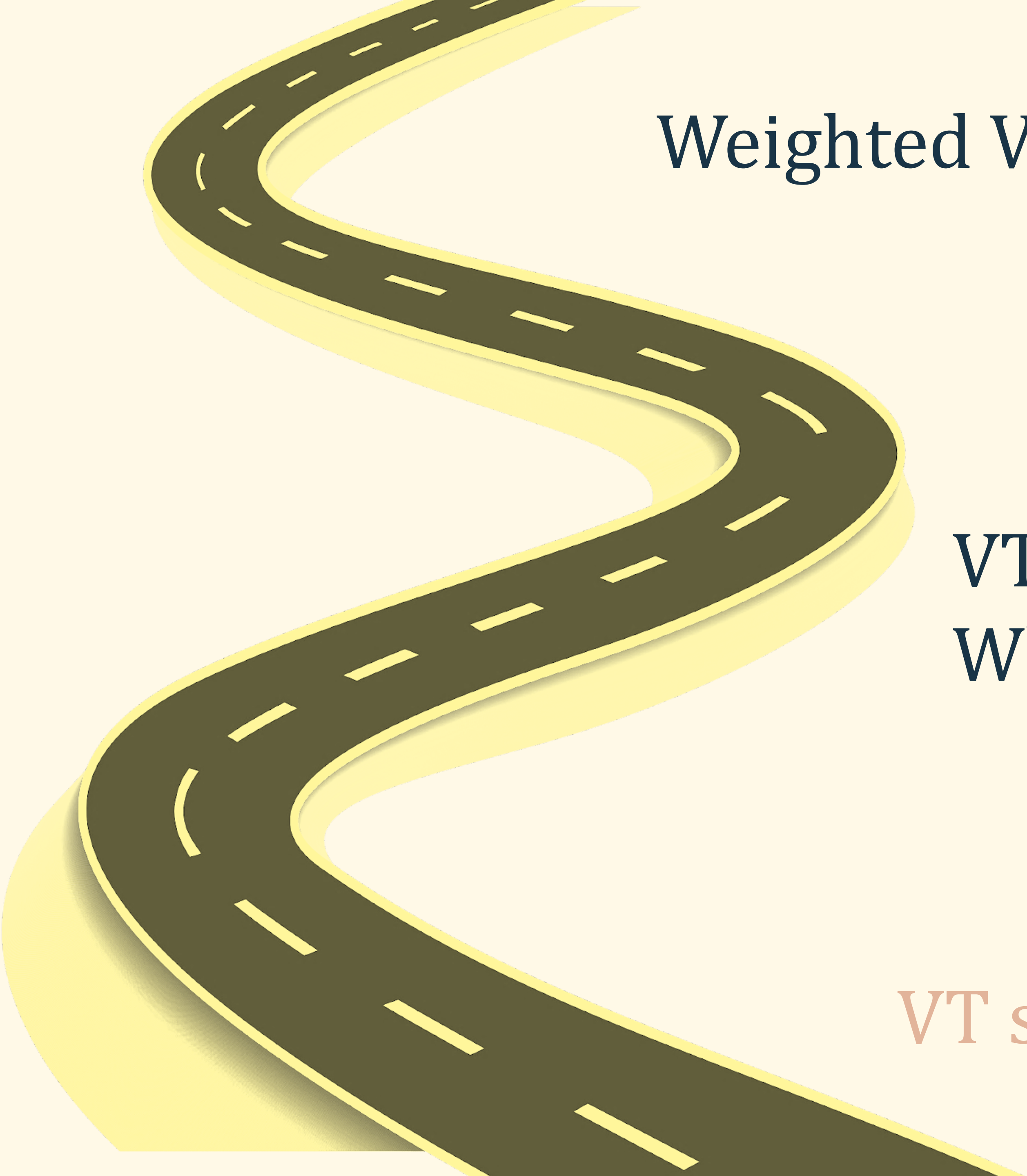
Alphabet	Error type	Redundancy
q	One del or one sub (edit error)	$\log n + O_q(\log\log n)$



Weighted VT sketch

VT sketch: ternary
Why it fails

VT sketch: binary



Weighted VT sketch

VT sketch: ternary
Why it fails

VT sketch: binary

Correcting one edit error: binary

VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

If substitution: $x = 0\ 0\ 1\ 0\ 1$ $\quad f(x) = 8$

[Lev 65, VT65]

VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

$$f(x) = 8$$

$$y = \ 0\ 0\ 1\ 1\ 1 \qquad f(y) = 12$$

$$f(y) - f(x) = \sum_{1}^{n} i \cdot (y_i - x_i) = e(y_i - x_i) = 4$$

Position of substitution

VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

$x = 0\ 0\ 1\ 0\ 1$ $\qquad f(x) = 8$

$y = 0\ 0\ 1\ 1\ 1$ $\qquad f(y) = 12$

Correcting one edit error: deletion

VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

$f(x) = 8$

$y = \ 0\ 0\ 0\ 1$ $\qquad f(y) = 4$

$$f(x) - f(y) = \sum_{d}^{n-1} y_i + d x_d = 4$$

VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

$$f(x) = 8$$

$$y = ?\,0\,0\,0\,1 \qquad f(y) = 4$$

$$f(x) - f(y) = \sum_{d}^{n-1} y_i + d x_d = 4$$

VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

$x = \ 0\ 0\ 1\ 0\ 1 \qquad f(x) = 8$

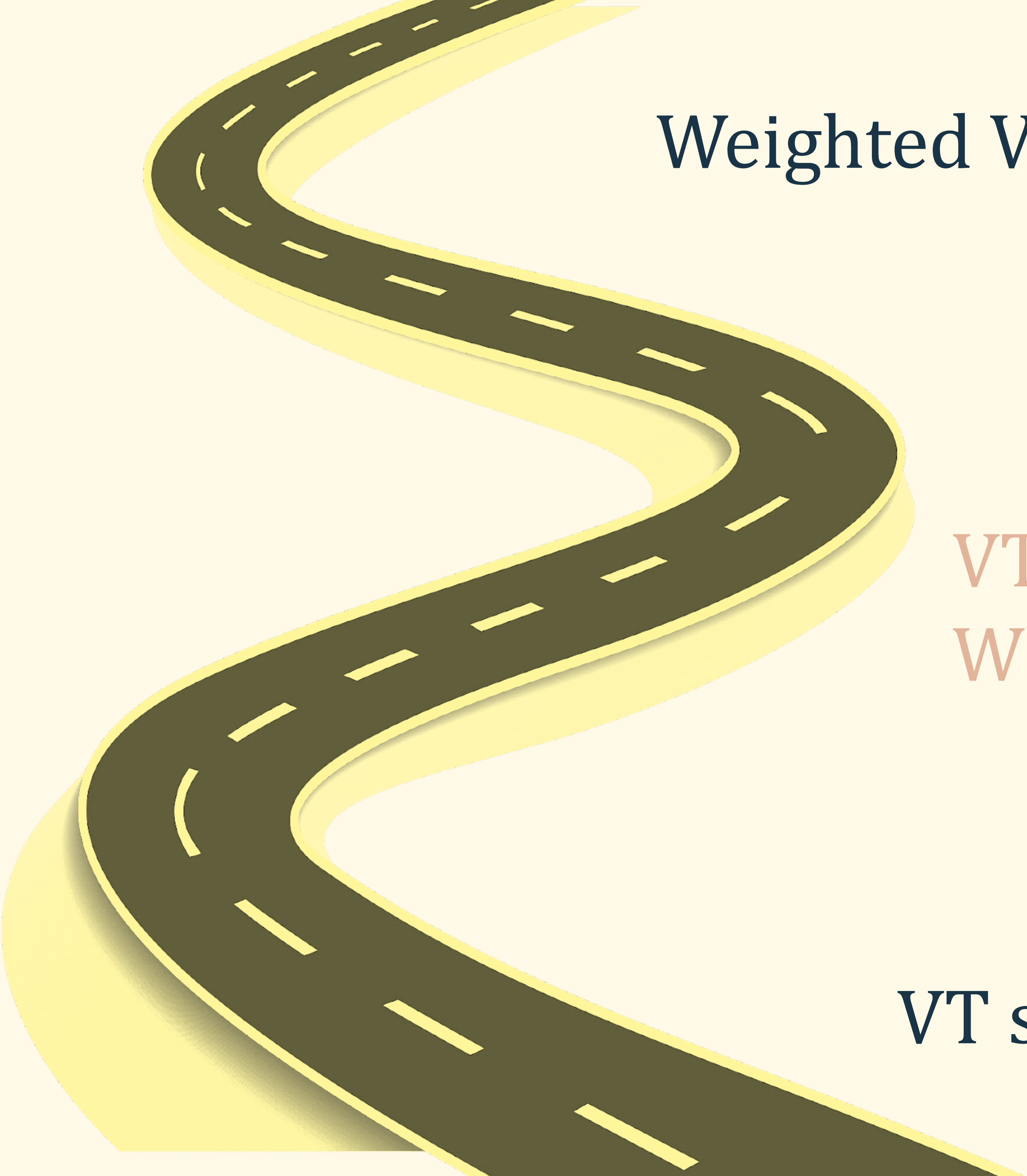
$y = \ 0\ 0\ 0\ ?\ 1 \qquad f(y) = 4$

$$f(x) - f(y) = \sum_{d}^{n-1} y_i + d x_d = 4$$

VT sketch

VT sketch: $f(x) = \sum_{i=1}^{n} i \cdot x_i$

$\log n + 2$ bits!



Weighted VT sketch

VT sketch: ternary
Why it fails

VT sketch: binary

VT code for ternary?

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$x = \ 0\ 2\ 1\ 0\ 1\ 2$ $\qquad f(x) = 24$

VT code for ternary: substitution

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$$f(x) = 24$$

$$y = \ 0\ 1\ 1\ 0\ 1\ 2$$

$$f(y) = 22$$

$$f(y) - f(x) = e(y_i - x_i) = 2$$

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$\#0 = 2$

$\#1 = 2$

$\#2 = 2$

$x = 2\ 1\ 1\ 0\ 1\ 2$

$x = 0\ 2\ 1\ 0\ 1\ 2$

$f(x) = 24$

$y = 0\ 1\ 1\ 0\ 1\ 2$

$f(y) = 22$

$$f(y) - f(x) = e(y_i - x_i) = 2$$

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$x = 0\ 2\ 1\ 0\ 1\ 2$ $\qquad f(x) = 24$

$y = 0\ 1\ 1\ 0\ 1\ 2$ $\qquad f(y) = 22$

VT code for ternary: deletion of 0

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$$f(x) = 24$$

$$y = \ 0\ 2\ 1\ 1\ 2$$

$$f(y) = 21$$

$$f(x) - f(y) = \sum_{d}^{n-1} y_i = 3$$

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$$x = 0\ 2\ 1\ 0\ 1\ 2 \qquad f(x) = 24$$

$$y = 0\ 2\ 1\ ?\ 1\ 2 \qquad f(y) = 21$$

$$f(x) - f(y) = \sum_{d}^{n-1} y_i = 3$$

Position of deletion

VT code for ternary: deletion of 1

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$$f(x) = 24$$

$$y = \; 0\ 2\ 0\ 1\ 2$$

$$f(y) = 18$$

$$f(x) - f(y) = \sum_{d}^{n-1} y_i + d = 6$$

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$x = 1\ 0\ 2\ 0\ 1\ 2$

$x = 0\ 2\ 1\ 0\ 1\ 2$ $\qquad f(x) = 24$

$y = 0\ 2\ 0\ 1\ 2$ $\qquad f(y) = 18$

$$f(x) - f(y) = \sum_{d}^{n-1} y_i + d = 6$$

$$x = 1\ 0\ 2\ 0\ 1\ 2$$

$$x = 0\ 2\ 1\ 0\ 1\ 2$$

$$\sum_{d}^{n-1} y_i + d$$

$$\sum_{d}^{n-1} y_i + d$$

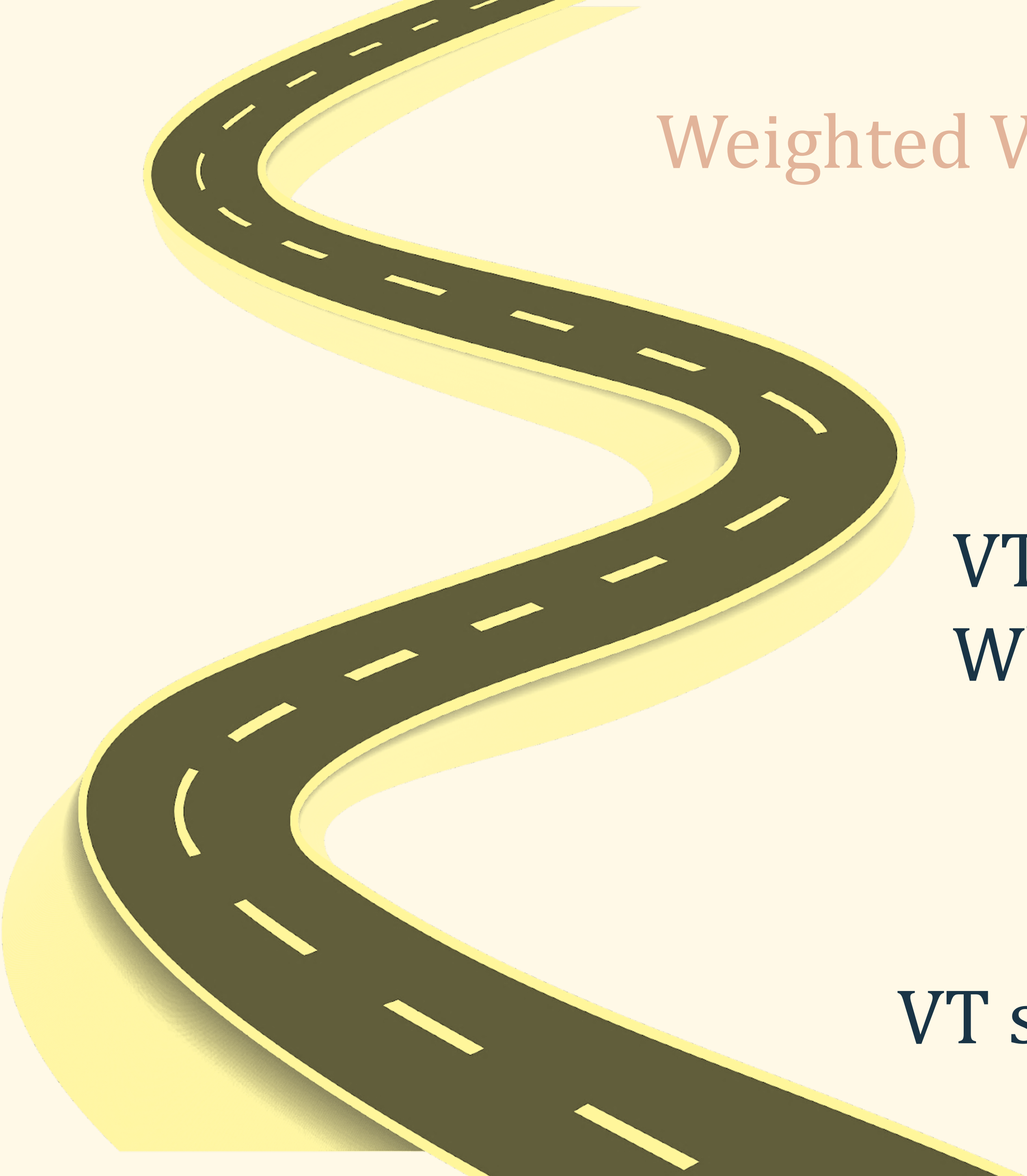
$x = 1\ 0\ 2\ 0\ 1\ 2$

$x = 0\ 2\ 1\ 0\ 1\ 2$

$$\sum_{d}^{n-1} y_i + d$$

$$\sum_{d}^{n-1} y_i + d$$

Fails when there is a chunk with an average 1



Weighted VT sketch

VT sketch: ternary
Why it fails

VT sketch: binary

Bias the weight!

$$f(x) = \sum_{i=1}^{n} i \cdot x_i$$

$$f(x) = \sum_{i=1}^{n} i \cdot w(x_i)$$

$$w(0) = 0$$
$$w(1) = 1$$
$$w(2) = 2 \log n$$

$$f(x) = \sum_{i=1}^{n} i \cdot w(x_i)$$

$$w(0) = 0$$

$$w(1) = 1$$

$$w(2) = 2 \log n$$

$$x = 1\ 0\ 2\ \ldots$$

$$f(x) = 1 + 6 \log n$$

$$x = 0\ 2\ 1\ \ldots$$

$$f(x) = 4 \log n + 3$$

As long as the chunk of avg 1 has length $< 2\log n$

$$x = 1\ \underbrace{0\,2\,0\,2\,0\,2\,0\,2\,0\,2}_{< 2\log n}\ 2\ \ldots$$

Run of 0 after deleting all 2's $< \log n$

$$x = 1\ \underbrace{0\ \ \ 0\ \ \ 0\ \ \ 0\ \ \ 0\ \ }_{< \log n}\ 2\ \ldots$$

Runlength replacement [SWGY17]

Weighted VT: encoding

x → Regular x'

↓

Weighted VT of x'

$\log n + \log\log n$ redundancy!

More in paper

1. Extends to general q
2. Binary code correcting one del and one sub
3. Binary code correcting one del and one adjacent trans