Edit Distance Cannot Be Computed in Strongly Subquadratic Time

(Unless SETH is false)

Presented by: Ethan Rambacher

Authors: Arturs Backurs and Piotr Indyk, MIT

STOC 2015

Edit Distance

- Minimum number of insertions, deletions, or substitutions of symbols to transform one string into another

```
TRAIN

 PAIL
```

Edit Distance

- Minimum number of insertions, deletions, or substitutions of symbols to transform one string into another

```
TRAIN
TRAIL (N->L)

 PAIL
```

Edit Distance

- Minimum number of insertions, deletions, or substitutions of symbols to transform one string into another

```
TRAIN
TRAIL
 RAIL (delete T)
 PAIL
```

Edit Distance

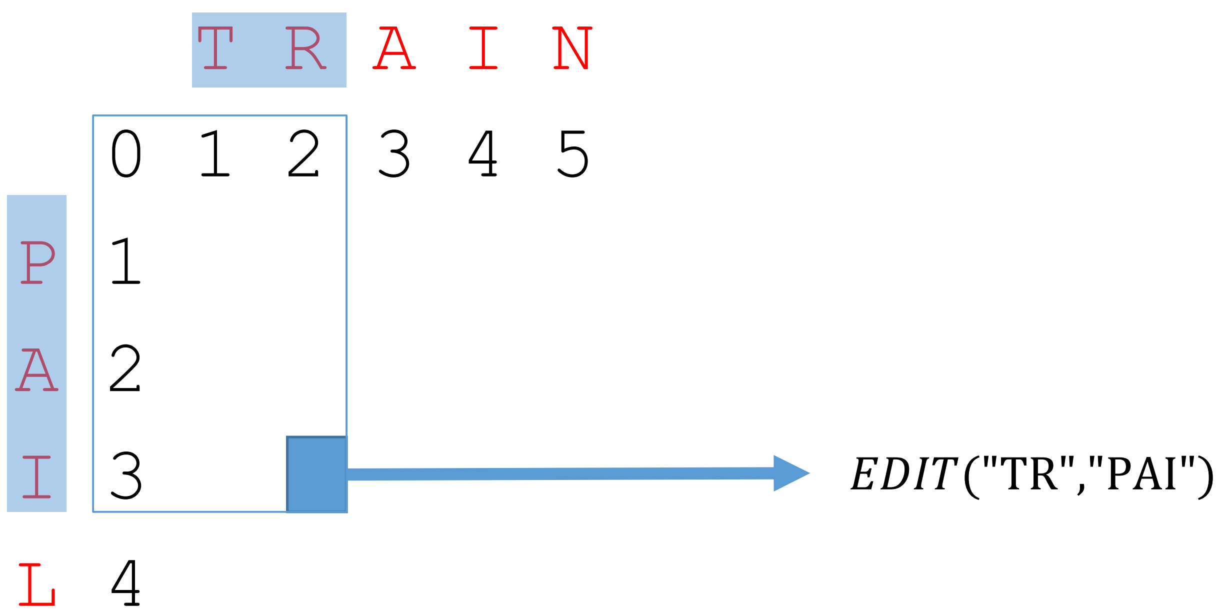
- Minimum number of insertions, deletions, or substitutions of symbols to transform one string into another

```
TRAIN
TRAIL
 RAIL
 PAIL (R->P)
```

EDIT(TRAIN,PAIL)=3

Computation of Edit Distance

- Dynamic programming technique yields $O(n^2)$ algorithm



Computation of Edit Distance

		T	R	A	I	N
	0	**1**	2	3	4	5
P	**1**	_				
A	2					
I	3					
L	4					

Computation of Edit Distance

		T	R	A	I	N
	0	**1**	**2**	3	4	5
P	1	**1**	_			
A	2					
I	3					
L	4					

Computation of Edit Distance

		T	R	A	I	N
	0	1	2	3	4	5
P	1	1	2	3	4	5
A	2	2	2	2	3	4
I	3	3	3	3	2	3
L	4	4	4	4	3	3

→ $EDIT(\text{"TRAIN"}, \text{"PAIL"}) = 3$

Computation of Edit Distance

- Best algorithm runs at $O\left(\frac{n^2}{\log^2 n}\right)$
- Can we achieve $O\left(n^{2-\delta}\right)$?

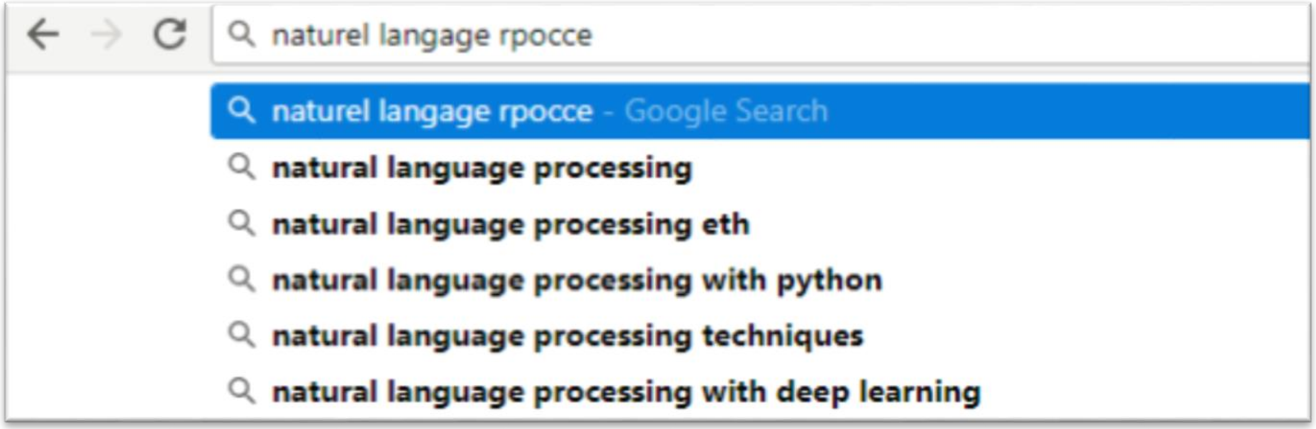
Applications of Edit Distance

- Computational biology

```
ATAGAGTACCCATAGATACAT TACG  AGCGCTAGC
ATT AGT CAGACCGA  C TGTACGTGACTGCT GC
```

Applications of Edit Distance

- Spell checking
- Natural language processing



Orthogonal Vectors Problem

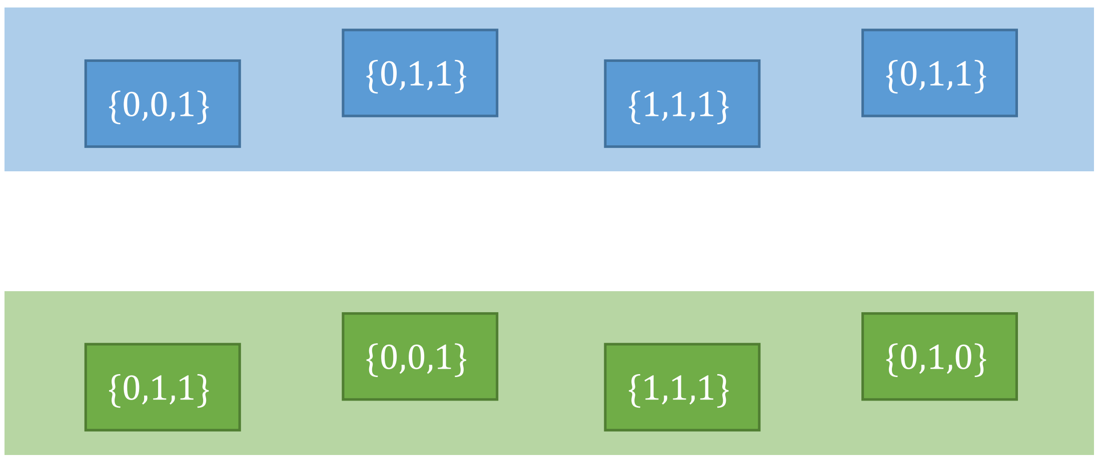
- Two sets $A, B \subseteq \{0,1\}^d$ such that $|A| = |B| = N$
- Is there a pair of orthogonal vectors?

$$x \in A, y \in B:$$

$$x \cdot y = \sum_{j=1}^{d} x_j y_j = 0$$

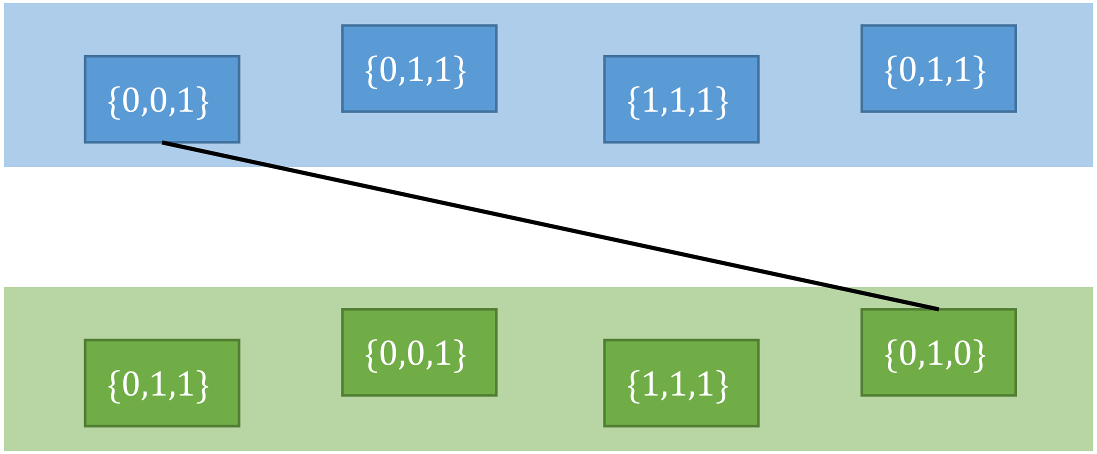
Orthogonal Vectors Problem

- Is there a pair of orthogonal vectors?



Orthogonal Vectors Problem

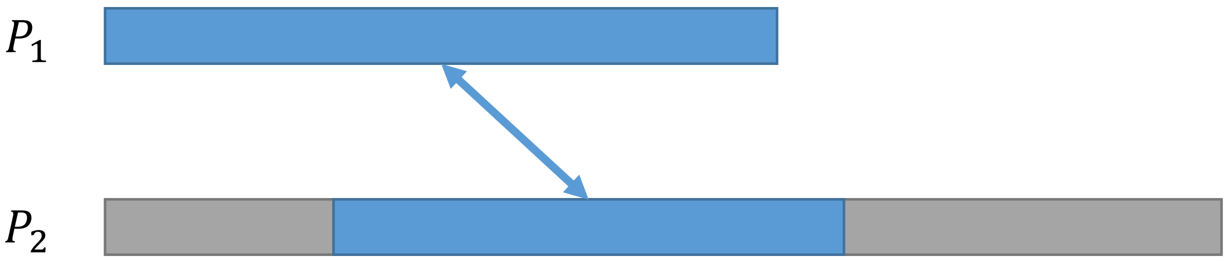
- Is there a pair of orthogonal vectors?



$$\{0,0,1\} \cdot \{0,1,0\} = (0 * 0) + (0 * 1) + (1 * 0) = 0$$

PATTERN

$$PATTERN(P_1, P_2) = \min_{\substack{x \text{ is a contiguous} \\ \text{subsequence of } P_2}} EDIT(P_1, x)$$



PATTERN Example

$$PATTERN(IRELAND, SWITZERLAND) = 3$$

SWITZ**ERLAND**

E**R**LAND

ELAND

IRELAND

PATTERN Example

$$PATTERN(STARSHIP, COMPUTERSCIENCE) = 4$$

COMPU**TERSCI**ENCE

TERSCI

T**A**RS**H**I

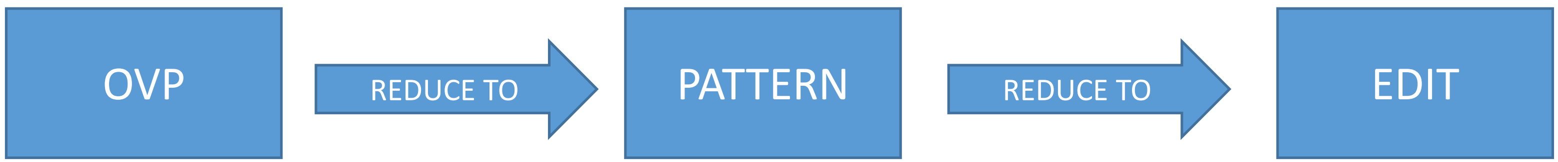
STARSHI**P**

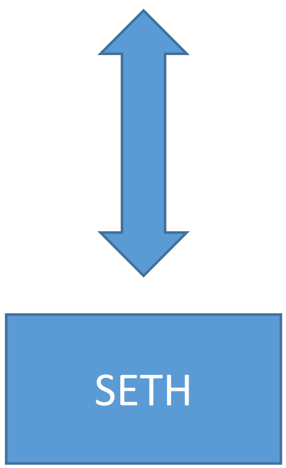
Strong Exponential Time Hypothesis (SETH)

SAT cannot be solved in subexponential time in the worst case

- Stronger than P != NP
  - If proven, SETH would imply P != NP

Proof Process



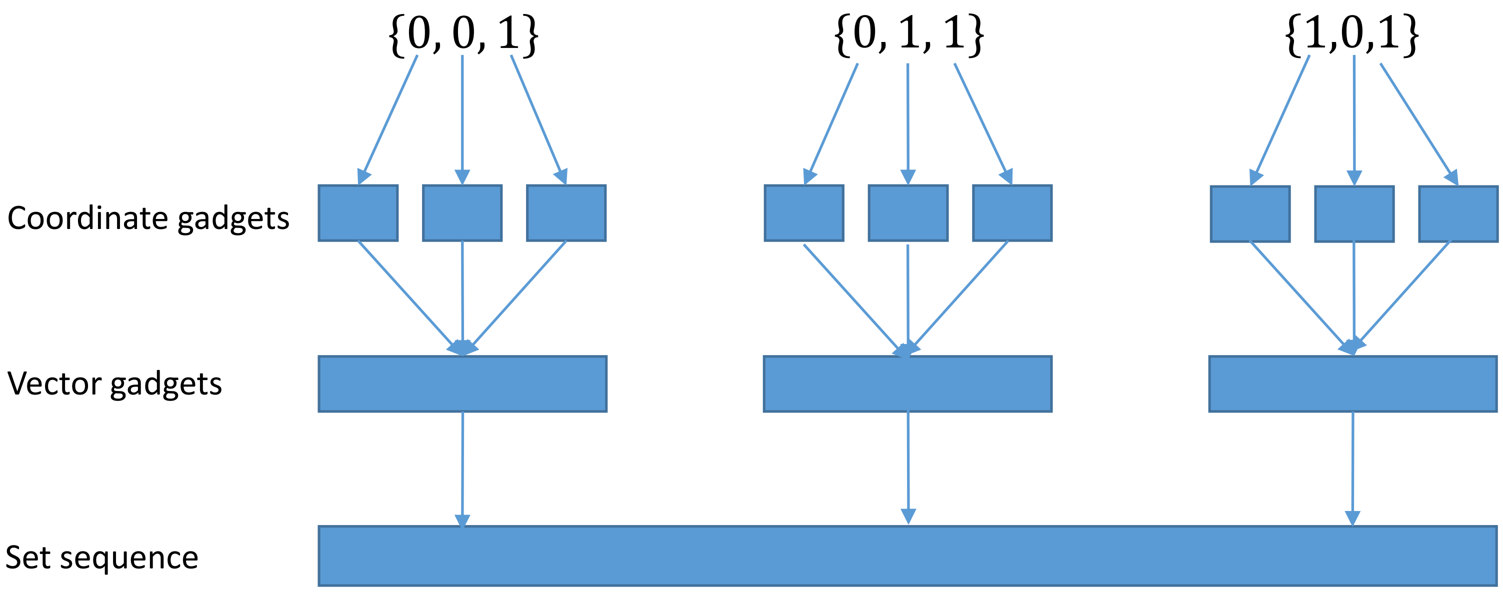


- $O\left(d^{O(1)} \cdot N^{2-\delta}\right)$ solution to OVP violates SETH
- Reduce OVP to EDIT

Intuition

- Convert each set of vectors into a string
- If there are orthogonal vectors, EDIT is “small”
- If there are no orthogonal vectors, EDIT is “large”

Intuition


{0, 0, 1}
{0, 1, 1}
{1,0,1}
Coordinate gadgets
Vector gadgets
Set sequence

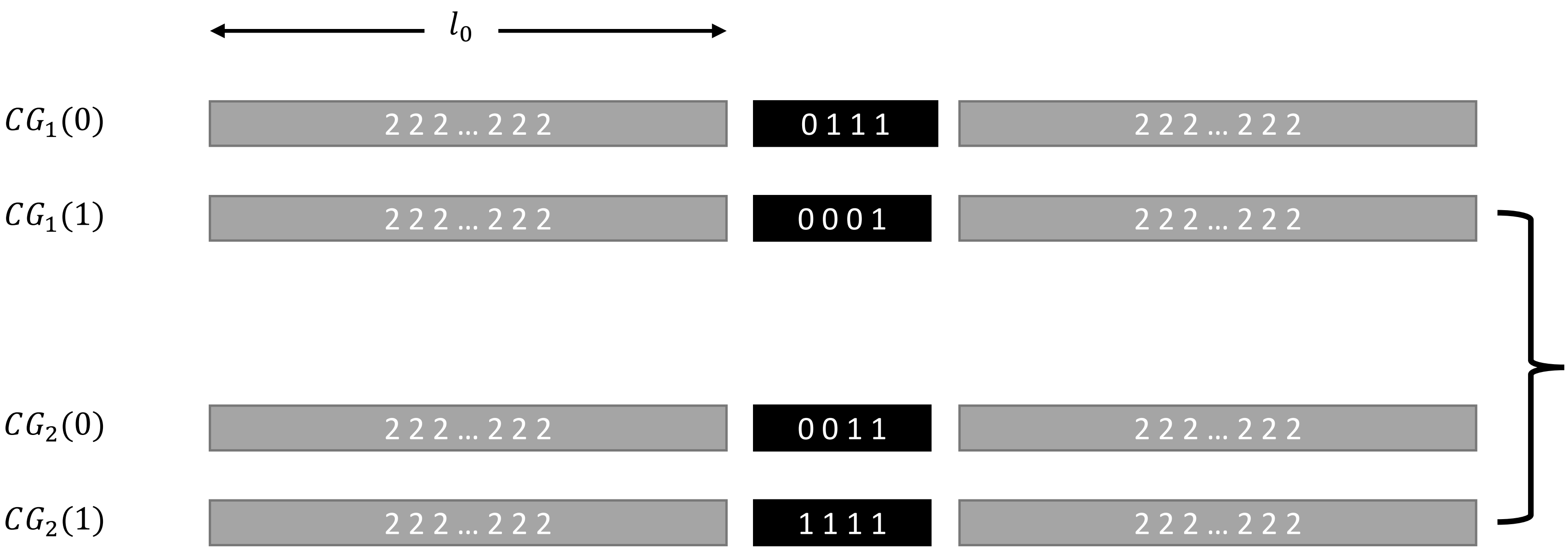
Coordinate Gadgets

$$CG_1(x) := \begin{cases} 2^{l_0}\ 0\ 1\ 1\ 1\ 2^{l_0} & if\ x = 0 \\ 2^{l_0}\ 0\ 0\ 0\ 1\ 2^{l_0} & if\ x = 1 \end{cases}$$

$$CG_2(x) := \begin{cases} 2^{l_0}\ 0\ 0\ 1\ 1\ 2^{l_0} & if\ x = 0 \\ 2^{l_0}\ 1\ 1\ 1\ 1\ 2^{l_0} & if\ x = 1 \end{cases}$$

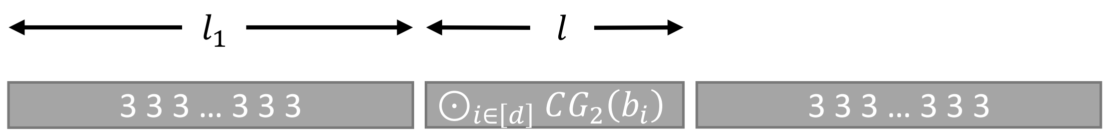
Thus, $EDIT\big(CG_1(x_1), CG_2(x_2)\big) = \begin{cases} 1 & if\ x_1 \cdot x_2 = 0 \\ 3 & if\ x_1 \cdot x_2 = 1 \end{cases}$

Visualization of the Coordinate Gadgets



Visualization of the Vector Gadgets

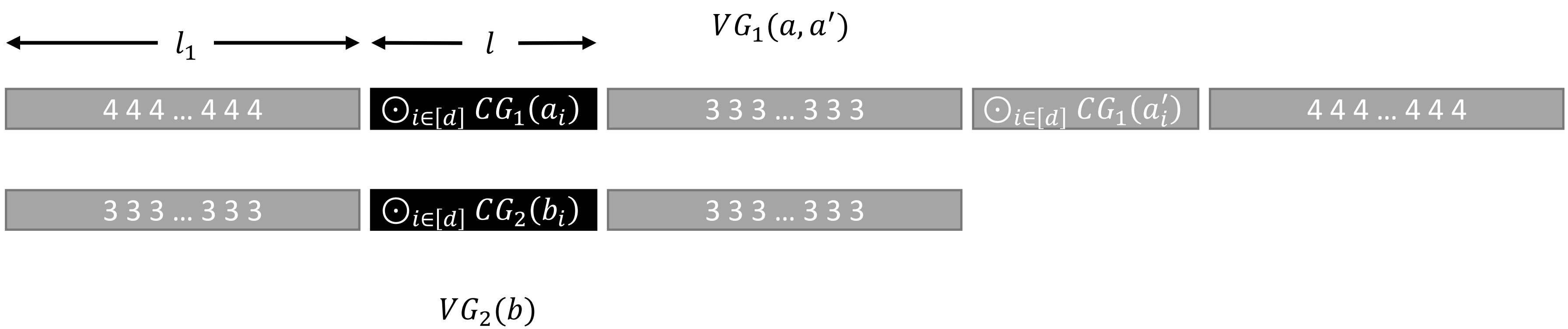
$VG_1(a, a')$



$VG_2(b)$

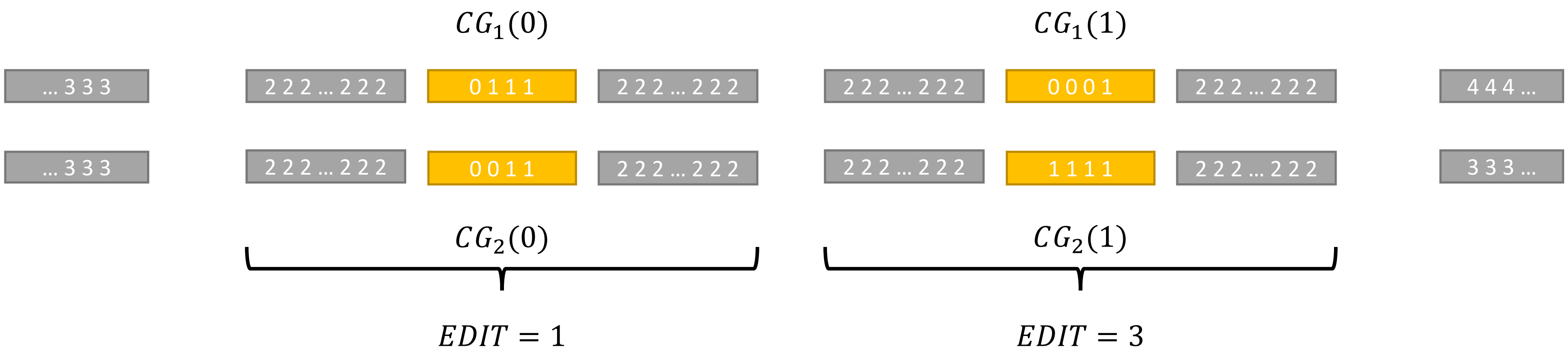
Let $l = \left|\bigodot_{i\in[d]} CG_1(a_i)\right| = d(4 + 2l_0)$

Vector Gadget Alignment – Case 1



1. Delete and substitute: $C' \coloneqq l_1 + l + l_1 = 2l_1 + d(2l_0 + 4)$
2. Transform concatenations of coordinate gadgets: $d + 2(a \cdot b)$
3. Total cost: $C' + d + 2(a \cdot b) = C + 2(a \cdot b)$

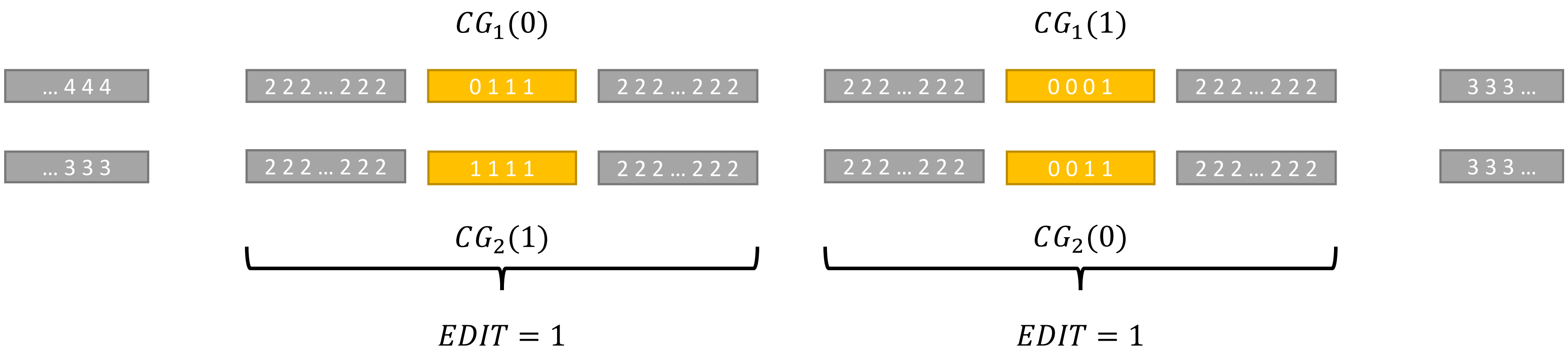
Transforming the Concatenations – Case 1.1



In this example we have $d = 2$:

$$a = \{0,1\}, b = \{0,1\}$$

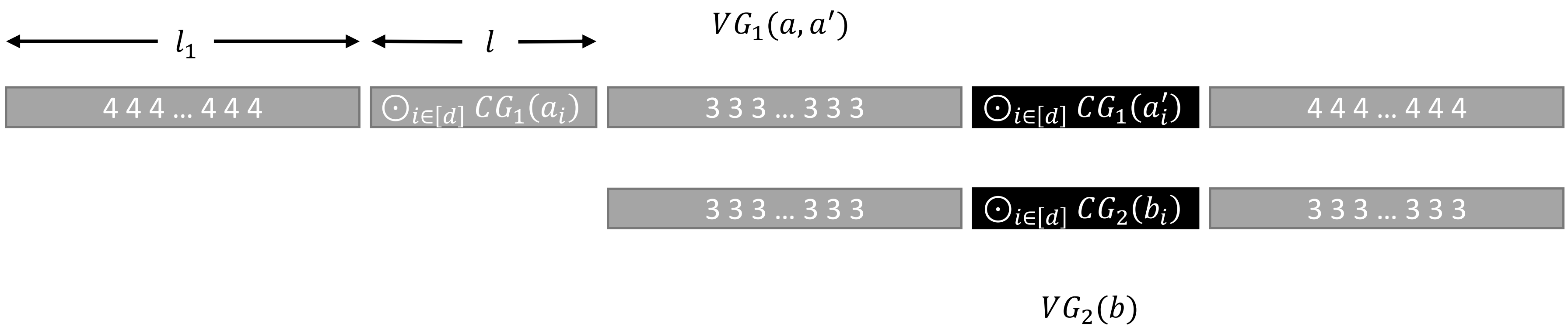
Transforming the Concatenations – Case 1.2



In this example we have $d = 2$:

$$a = \{0,1\}, b = \{1,0\}$$

Vector Gadget Alignment – Case 2



1. Delete and substitute: $C' \coloneqq l_1 + l + l_1 = 2l_1 + d(2l_0 + 4)$
2. Transform concatenations of coordinate gadgets: $d + 2(a' \cdot b)$
3. Total cost: $C' + d + 2(a' \cdot b) = C + 2(a' \cdot b)$

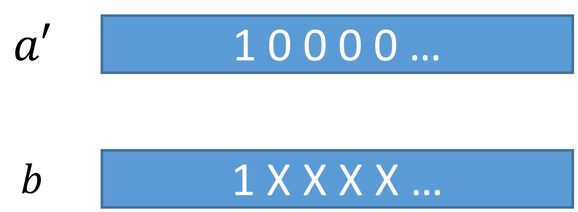
Edit distance of Vector Gadgets

$$EDIT\left(VG_1(a, a'), VG_2(b)\right) = K + 2 \cdot \min(a \cdot b, a' \cdot b) \coloneqq X$$

We want to get rid of the “min” part

Edit distance of Vector Gadgets

- Let $a' = 1\,0^{d-1}$
- Let b begin with 1 (i.e. $b_1 = 1$)
- Then, $a' \cdot b = 1$
- This removes the “min” part of our formula



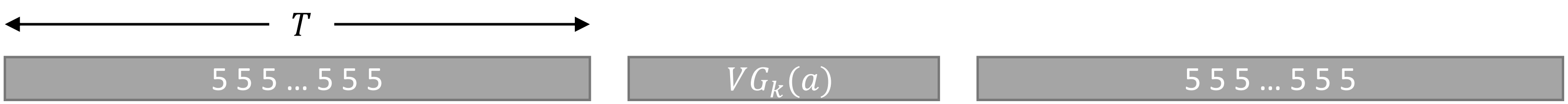
Edit distance of Vector Gadgets

$$EDIT\big(VG_1(a), VG_2(b)\big) = \begin{cases} E_s & if\ a \cdot b = 0 \\ E_u := E_s + 2 & if\ a \cdot b \geq 1 \end{cases}$$

Now there is a threshold between orthogonal vectors and non-orthogonal vectors

New Vector Gadgets

- Let t be the maximum of the lengths of VG_1, VG_2
- Let $T = 1000d \cdot t = \Theta(d^3)$



$$VG'_k(a)$$

Set Sequences

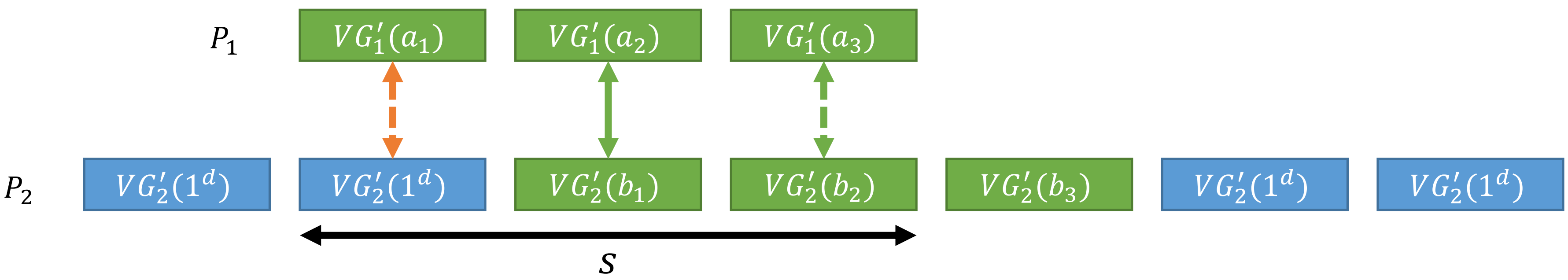
$$P_1 = \bigodot_{a \in A} VG_1'(a)$$

$VG_1'(a_1)$ $VG_1'(a_2)$ $VG_1'(a_3)$

$VG_2'(1^d)$ $VG_2'(1^d)$ $VG_2'(b_1)$ $VG_2'(b_2)$ $VG_2'(b_3)$ $VG_2'(1^d)$ $VG_2'(1^d)$

$$P_2 = \left(\bigodot_{i=1}^{|A|-1} VG_2'(1^d)\right)\left(\bigodot_{b \in B} VG_2'(b)\right)\left(\bigodot_{i=1}^{|A|-1} VG_2'(1^d)\right)$$

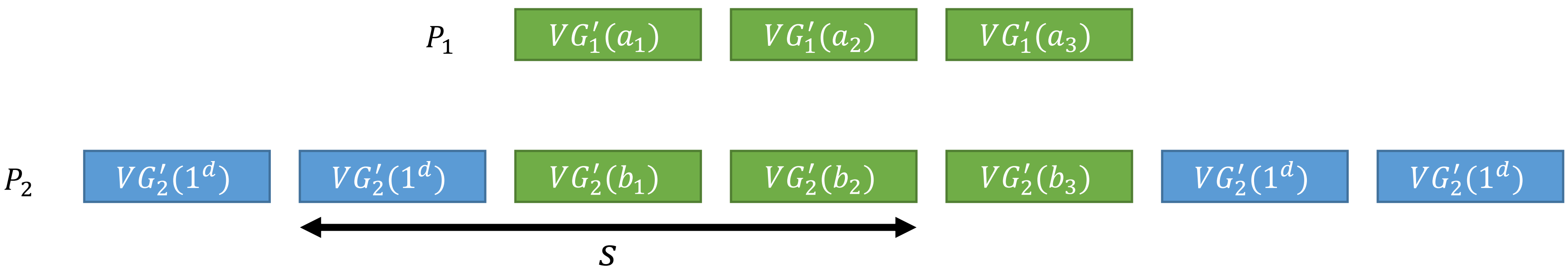
PATTERN distance with orthogonal vectors



Sequence s made of $|A|$ vector gadgets in P_2
We have **at least one** alignment

$$PATTERN(P_1, P_2) \leq E_u \cdot |A| - 2 = X - 2$$

PATTERN distance with no orthogonal vectors

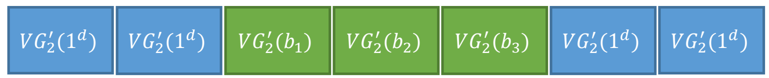


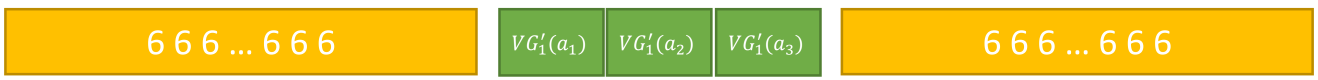
Sequence s made of $|A|$ vector gadgets in P_2
We have no alignments; each VG transformation costs E_u

$$PATTERN(P_1, P_2) = E_u \cdot |A| \coloneqq X$$

Reducing PATTERN to EDIT

$$P_2' := P_2$$

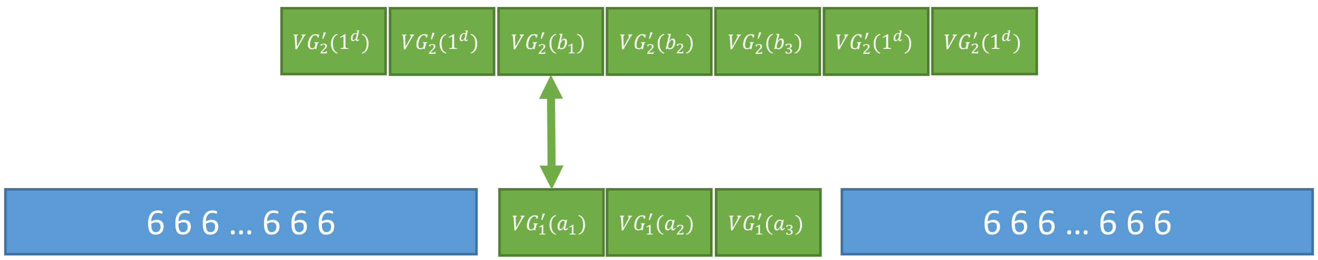




$$P_1' := 6^{|P_2'|}\, P_1\, 6^{|P_2'|}$$

Case 1: Orthogonal vectors

$$P_2' := P_2$$



$$P_1' := 6^{|P_2'|}\ P_1\ 6^{|P_2'|}$$

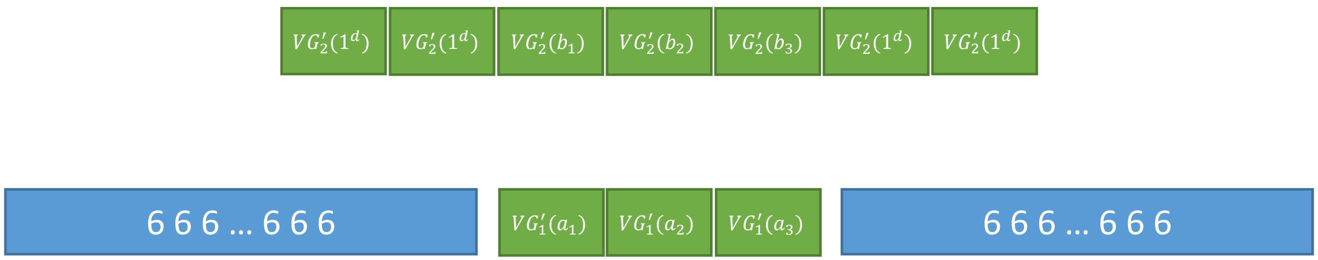
Green: $\leq X - 2$ transformations

Blue: $\leq 2 \cdot |P_2'|$ transformations

$$EDIT(P_1', P_2') = 2 \cdot |P_2'| + X - 2 = \mathrm{Y} - 2$$

Case 2: No orthogonal vectors

$$P_2' := P_2$$



$$P_1' := 6^{|P_2'|}\ P_1\ 6^{|P_2'|}$$

Green: X transformations

Blue: $2 \cdot |P_2'|$ transformations

$$\mathrm{EDIT}(\mathrm{P}_1', P_2') = 2 \cdot |P_2'| + X = Y$$

Reducing PATTERN to EDIT

- Therefore:
  - If there is a pair of orthogonal vectors:
$$EDIT(P_1', P_2') \leq Y - 2$$
  - If there is no pair of orthogonal vectors:
$$EDIT(P_1', P_2') = Y$$

We now have a threshold for solving OVP.

Conclusion

- If EDIT can be computed in $O\left(n^{2-\delta}\right)$ then OVP can be solved in $d^{O(1)} \cdot N^{2-\delta}$
- If this is true, SETH must be false