Faster Pattern Matching under Edit Distance

Panagiotis Charalampopoulos\textsuperscript{1}, Tomasz Kociumaka\textsuperscript{2}, Philip Wellnitz\textsuperscript{2}

1. Birkbeck, University of London, UK
2. Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany

FOCS 2022

Denver, USA
The Problem

Pattern Matching
Given a text $T$ and a pattern $P$, compute the occurrences of $P$ in $T$.

Pattern Matching under Edit Distance
Given a text $T$, a pattern $P$, and an integer threshold $k$, compute the (starting positions of) substrings of $T$ that are at edit distance at most $k$ from $P$.

P. Charalampopoulos, T. Kociumaka, P. Wellnitz
Faster Pattern Matching under Edit Distance
The Problem

**Pattern Matching**

Given a text $T$ and a pattern $P$, compute the occurrences of $P$ in $T$. 
The Problem

Pattern Matching

Given a text $T$ and a pattern $P$, compute the occurrences of $P$ in $T$. 

\[
\begin{array}{c}
T \\
p a n c a k e \\
\hline
P \\
c a k e
\end{array}
\]
The Problem

Pattern Matching
Given a text $T$ and a pattern $P$, compute the occurrences of $P$ in $T$.

Pattern Matching under Edit Distance
Given a text $T$, a pattern $P$, and an integer threshold $k$, compute the (starting positions of) substrings of $T$ that are at edit distance at most $k$ from $P$. 
The Problem

Pattern Matching

Given a text $T$ and a pattern $P$, compute the occurrences of $P$ in $T$.

Pattern Matching under Edit Distance

Given a text $T$, a pattern $P$, and an integer threshold $k$, compute the (starting positions of) substrings of $T$ that are at edit distance at most $k$ from $P$. 
History and our Result
History and our Result

\(\mathcal{O}(n^2)\)  [Sellers; J. Algorithms 1980]
History and our Result

$O(n^2)$  [Sellers; J. Algorithms 1980]

$O(nk^2)$  [Landau, Vishkin; JCSS 1988]
History and our Result

\[ \mathcal{O}(n^2) \] [Sellers; J. Algorithms 1980]

\[ \mathcal{O}(nk^2) \] [Landau, Vishkin; JCSS 1988]

\[ \mathcal{O}(nk) \] [Landau, Vishkin; J. Algorithms 1989]
### History and our Result

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{O}(n^2) )</td>
<td>[Sellers; J. Algorithms 1980]</td>
</tr>
<tr>
<td>( \mathcal{O}(nk^2) )</td>
<td>[Landau, Vishkin; JCSS 1988]</td>
</tr>
<tr>
<td>( \mathcal{O}(nk) )</td>
<td>[Landau, Vishkin; J. Algorithms 1989]</td>
</tr>
<tr>
<td>( \tilde{\mathcal{O}}(n + k^{8+1/3} \cdot n/m^{1/3}) )</td>
<td>[Sahinalp, Vishkin; FOCS 1996]</td>
</tr>
<tr>
<td>( \Omega(k^2) )</td>
<td>[Backurs, Indyk; SICOMP 2018]</td>
</tr>
</tbody>
</table>
## History and our Result

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Authors and Journals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{O}(n^2)$</td>
<td>Sellers; J. Algorithms 1980</td>
</tr>
<tr>
<td>$\mathcal{O}(nk^2)$</td>
<td>Landau, Vishkin; JCSS 1988</td>
</tr>
<tr>
<td>$\mathcal{O}(nk)$</td>
<td>Landau, Vishkin; J. Algorithms 1989</td>
</tr>
<tr>
<td>$\tilde{\mathcal{O}}(n + k^{8+1/3} \cdot n/m^{1/3})$</td>
<td>Sahinalp, Vishkin; FOCS 1996</td>
</tr>
<tr>
<td>$\mathcal{O}(n + k^4 \cdot n/m)$</td>
<td>Cole, Hariharan; SICOMP 2002</td>
</tr>
</tbody>
</table>
History and our Result

\[ \mathcal{O}(n^2) \] [Sellers; J. Algorithms 1980]

\[ \mathcal{O}(nk^2) \] [Landau, Vishkin; JCSS 1988]

\[ \mathcal{O}(nk) \] [Landau, Vishkin; J. Algorithms 1989]

\[ \tilde{\mathcal{O}}(n + k^{8+1/3} \cdot n/m^{1/3}) \] [Sahinalp, Vishkin; FOCS 1996]

\[ \mathcal{O}(n + k^4 \cdot n/m) \] [Cole, Hariharan; SICOMP 2002]

\[ \tilde{\mathcal{O}}(n + k^{3.5} \cdot n/m) \] This work
History and our Result

\[ \mathcal{O}(n^2) \] [Sellers; J. Algorithms 1980]

\[ \mathcal{O}(nk^2) \] [Landau, Vishkin; JCSS 1988]

\[ \mathcal{O}(nk) \] [Landau, Vishkin; J. Algorithms 1989]

\[ \tilde{\mathcal{O}}(n + k^{8+1/3} \cdot n/m^{1/3}) \] [Sahinalp, Vishkin; FOCS 1996]

\[ \mathcal{O}(n + k^4 \cdot n/m) \] [Cole, Hariharan; SICOMP 2002]

\[ \tilde{\mathcal{O}}(n + k^{3.5} \cdot n/m) \] This work

\[ \Omega(k^2) \] [Backurs, Indyk; SICOMP 2018]
History and our Result

\[ t(n, k) \approx n^2 \]

\[ t(n, k) \approx n^{4/3} \]

\[ t(n, k) \approx n \]

- \( k \approx 1 \)
- \( k \approx n^{1/4} \)
- \( k \approx n^{1/3} \)
- \( k \approx n^{1/2} \)
- \( k \approx n \)

- \( O(nk), LV \)
- \( O(n + k^4), CH \)
- \( \Omega(k^2), BI \)
History and our Result

\[ t(n, k) \approx n^2 \]

\[ t(n, k) \approx n^{7/5} \]

\[ t(n, k) \approx n^{4/3} \]

\[ t(n, k) \approx n \]

\[ k \approx 1 \]

\[ k \approx n^{1/4} \]

\[ k \approx n^{2/7} \]

\[ k \approx n^{1/3} \]

\[ k \approx n^{2/5} \]

\[ k \approx n^{1/2} \]

\[ k \approx n \]

\[ O(nk), LV \]

\[ \tilde{O}(n + k^{3.5}), CH \]

\[ O(n + k^4), This\ work \]

\[ \Omega(k^2), BI \]

P. Charalampopoulos, T. Kociumaka, P. Wellnitz

Faster Pattern Matching under Edit Distance
The Structure of Pattern Matching

Fact [folklore]
Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:

- The pattern $P$ has at most one occurrence in $T$.
- The pattern $P$ is periodic.
Fact [folklore] Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:
**The Structure of Pattern Matching**

**Fact [folklore]** Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:

- The pattern $P$ has at most one occurrence in $T$. 

![Diagram of pattern and text](image)
Fact [folklore] Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:

- The pattern $P$ has at most one occurrence in $T$.
- The pattern $P$ is periodic.
The Structure of Pattern Matching

**Fact [folklore]** Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:

- The pattern $P$ has at most one occurrence in $T$.
- The pattern $P$ is periodic.
Fact [folklore] Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:

- The pattern $P$ has at most one occurrence in $T$.
- The pattern $P$ is periodic.
The Structure of Pattern Matching

Fact [folklore] Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:

- The pattern $P$ has at most one occurrence in $T$.
- The pattern $P$ is periodic.
**Fact [folklore]** Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:

- The pattern $P$ has at most one occurrence in $T$.
- The pattern $P$ is periodic.
Fact [folklore] Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$ at least one of the following holds:

- The pattern $P$ has at most one occurrence in $T$.
- The pattern $P$ is periodic.

The fragment of $T$ spanned by $P$’s occurrences is periodic as well.
Theorem [CKW; FOCS'20]

Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$, and a threshold $k \leq m$ at least one of the following holds:

- The pattern $P$ has $O(k^2 k^{-})$ error occurrences in $T$.
- $P$ is almost periodic: at edit distance $< 2k$ from a string with period $O(m/k)$.

We call this a tile decomposition of $P$ with respect to $Q$. It does not match any of its rotations.
The Structure of Pattern Matching under Edit Distance

**Theorem [CKW; FOCS’20]** Given a pattern \( P \) of length \( m \) and a text \( T \) of length \( n \leq \frac{3}{2} m \), and a threshold \( k \leq m \) at least one of the following holds:

- The pattern \( P \) has \( O(k^2) \) \( k \)-error occurrences in \( T \).
- The pattern is almost periodic: at edit distance \( < 2k \) from a string with period \( O(m/k) \).
**The Structure of Pattern Matching under Edit Distance**

**Theorem [CKW; FOCS’20]** Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$, and a threshold $k \leq m$ at least one of the following holds:

- The pattern $P$ has $\mathcal{O}(k^2)$ $k$-error occurrences in $T$. 

$P$ will denote a primitive string; it does not match any of its rotations. We call this a tile decomposition of $P$ with respect to $Q$. 

P. Charalampopoulos, T. Kociumaka, P. Wellnitz

Faster Pattern Matching under Edit Distance
The Structure of Pattern Matching under Edit Distance

**Theorem [CKW; FOCS’20]** Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$, and a threshold $k \leq m$ at least one of the following holds:

- The pattern $P$ has $\mathcal{O}(k^2)$ $k$-error occurrences in $T$.
- The pattern is almost periodic: at edit distance $< 2k$ from a string with period $\mathcal{O}(m/k)$.  

We call this a tile decomposition of $P$ with respect to $Q$. 

P. Charalampopoulos, T. Kociumaka, P. Wellnitz
Faster Pattern Matching under Edit Distance
The Structure of Pattern Matching under Edit Distance

Theorem [CKW; FOCS’20] Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2}m$, and a threshold $k \leq m$ at least one of the following holds:

- The pattern $P$ has $O(k^2)$ $k$-error occurrences in $T$.
- The pattern is almost periodic: at edit distance $< 2k$ from a string with period $O(m/k)$. This is the bottleneck.
The Structure of Pattern Matching under Edit Distance

**Theorem [CKW; FOCS’20]** Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$, and a threshold $k \leq m$ at least one of the following holds:

- The pattern $P$ has $O(k^2)$ $k$-error occurrences in $T$.
- The pattern is almost periodic: at edit distance $< 2k$ from a string with period $O(m/k)$. **This is the bottleneck.**

\[ P \]
\[ Q^{\infty} \]

\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ Q \]
\[ \ldots \]

$Q$ will denote a **primitive string**; it does not match any of its rotations.
Theorem [CKW; FOCS’20] Given a pattern $P$ of length $m$ and a text $T$ of length $n \leq \frac{3}{2} m$, and a threshold $k \leq m$ at least one of the following holds:

- The pattern $P$ has $O(k^2)$ $k$-error occurrences in $T$.
- The pattern is almost periodic: at edit distance $< 2k$ from a string with period $O(m/k)$. This is the bottleneck.

$Q^\infty$ will denote a primitive string; it does not match any of its rotations.

We call this a tile decomposition of $P$ with respect to $Q$. 
The PILLAR Model and the Reduction of [CKW’20]

For any setting, e.g., when the strings are given in compressed form, an efficient implementation of the primitive operations yields a fast algorithm. Standard setting: The primitive operations take $O(1)$ time after an $O(n)$-time preprocessing. $O(k^4 \cdot n/m)$ PILLAR-time algorithm [CKW’20] matches [Cole, Hariharan; SICOMP 2002] for the standard setting.

Reduction [CKW’20]: An algorithm that solves the almost periodic case in $\tilde{O}(k^a \cdot n/m)$ PILLAR-time, for $a \geq 3$, implies an algorithm that solves the general case in $\tilde{O}(k^a \cdot n/m)$ PILLAR-time.
The PILLAR Model and the Reduction of [CKW’20]

In the PILLAR model [CKW’20], algorithms rely on primitive operations.
In the PILLAR model [CKW’20], algorithms rely on primitive operations.

For any setting, e.g., when the strings are given in compressed form, an efficient implementation of the primitive operations yields a fast algorithm.
The PILLAR Model and the Reduction of [CKW’20]

In the PILLAR model [CKW’20], algorithms rely on primitive operations.

For any setting, e.g., when the strings are given in compressed form, an efficient implementation of the primitive operations yields a fast algorithm.

Standard setting: The primitive operations take $O(1)$ time after an $O(n)$-time preprocessing.
The PILLAR Model and the Reduction of [CKW’20]

In the PILLAR model [CKW’20], algorithms rely on primitive operations.

For any setting, e.g., when the strings are given in compressed form, an efficient implementation of the primitive operations yields a fast algorithm.

Standard setting: The primitive operations take $O(1)$ time after an $O(n)$-time preprocessing.

$O(k^4 \cdot n/m)$ PILLAR-time algorithm [CKW’20] matches [Cole, Hariharan; SICOMP 2002] for the standard setting.
In the PILLAR model [CKW’20], algorithms rely on primitive operations. For any setting, e.g., when the strings are given in compressed form, an efficient implementation of the primitive operations yields a fast algorithm.

**Standard setting:** The primitive operations take $\mathcal{O}(1)$ time after an $\mathcal{O}(n)$-time preprocessing.

$\mathcal{O}(k^4 \cdot n/m)$ PILLAR-time algorithm [CKW’20] matches [Cole, Hariharan; SICOMP 2002] for the standard setting.

**Reduction [CKW’20]:** An algorithm that solves the almost periodic case in $\tilde{\mathcal{O}}(k^a \cdot n/m)$ PILLAR-time, for $a \geq 3$, implies an algorithm that solves the general case in $\tilde{\mathcal{O}}(k^a \cdot n/m)$ PILLAR-time.
Dynamic Puzzle Matching

Input: An integer

Maintain: A sequence $I = (U_1, V_1) \cdots (U_z, V_z)$ of pairs from $F^2$.

Updates: Insertions and deletions of pairs in $I$.

Queries: Compute the $k$-error occurrences of $U_1 \cdots U_z$ in $V_1 \cdots V_z$.

After $\tilde{O}(k^3)$-time preprocessing, updates and queries take $\tilde{O}(k)$ time.
Dynamic Puzzle Matching

**Input:** An integer $k$ and a family $\mathcal{F}$ of strings containing a distinguished primitive string $Q$ with $\sum_{F \in \mathcal{F}} \delta_E(F, Q) = \mathcal{O}(k)$. 
**Dynamic Puzzle Matching**

**Input:** An integer $k$ and a family $\mathcal{F}$ of strings containing a distinguished primitive string $Q$ with $\sum_{F \in \mathcal{F}} \delta_E(F, Q) = O(k)$.

**Maintain:** A sequence $\mathcal{I} = (U_1, V_1) \cdots (U_z, V_z)$ of pairs from $\mathcal{F}^2$.

- Updates: Insertions and deletions of pairs in $\mathcal{I}$.
- Queries: Compute the $k$-error occurrences of $U_1 \cdots U_z$ in $V_1 \cdots V_z$.

After $\tilde{\Theta}(k^3)$-time preprocessing, updates and queries take $\tilde{\Theta}(k)$ time.
Input: An integer $k$ and a family $\mathcal{F}$ of strings containing a distinguished primitive string $Q$ with $\sum_{F \in \mathcal{F}} \delta_E(F, Q) = \mathcal{O}(k)$.

Maintain: A sequence $\mathcal{I} = (U_1, V_1) \cdots (U_z, V_z)$ of pairs from $\mathcal{F}^2$.

Updates: Insertions and deletions of pairs in $\mathcal{I}$. 
Dynamic Puzzle Matching

**Input:** An integer $k$ and a family $\mathcal{F}$ of strings containing a distinguished primitive string $Q$ with $\sum_{F \in \mathcal{F}} \delta_E(F, Q) = \mathcal{O}(k)$.

**Maintain:** A sequence $\mathcal{I} = (U_1, V_1) \cdots (U_z, V_z)$ of pairs from $\mathcal{F}^2$.

**Updates:** Insertions and deletions of pairs in $\mathcal{I}$.

**Queries:** Compute the $k$-error occurrences of $U_1 \cdots U_z$ in $V_1 \cdots V_z$. 
**Dynamic Puzzle Matching**

**Input:** An integer $k$ and a family $\mathcal{F}$ of strings containing a distinguished primitive string $Q$ with $\sum_{F \in \mathcal{F}} \delta_E(F, Q) = \mathcal{O}(k)$.

**Maintain:** A sequence $\mathcal{I} = (U_1, V_1) \cdots (U_z, V_z)$ of pairs from $\mathcal{F}^2$.

**Updates:** Insertions and deletions of pairs in $\mathcal{I}$.

**Queries:** Compute the $k$-error occurrences of $U_1 \cdots U_z$ in $V_1 \cdots V_z$.

After $\tilde{\mathcal{O}}(k^3)$-time preprocessing, updates and queries take $\tilde{\mathcal{O}}(k)$ time.
Using Dynamic Puzzle Matching

Think of: $k = 4$ and $|Q| \approx \sqrt{m}$. 

P. Charalampopoulos, T. Kociumaka, P. Wellnitz

Faster Pattern Matching under Edit Distance
Using Dynamic Puzzle Matching

Think of: \( k = 4 \) and \( |Q| \approx \sqrt{m} \).

Each string has \( \mathcal{O}(k) \) special tiles.
Using Dynamic Puzzle Matching

Think of: $k = 4$ and $|Q| \approx \sqrt{m}$. 

\[ P \]
\[ T \]
Using Dynamic Puzzle Matching

Think of: \( k = 4 \) and \( |Q| \approx \sqrt{m} \).

> \( k \) copies of \( Q \) in \( P \) \( \implies \geq 1 \) must be matched exactly.
Using Dynamic Puzzle Matching

Think of: $k = 4$ and $|Q| \approx \sqrt{m}$.

$k$ copies of $Q$ in $P \implies \geq 1$ must be matched exactly

Starting positions of $k$-error occs in $T$ are within $O(k)$ from endpoints of tiles.
Think of: $k = 4$ and $|Q| \approx \sqrt{m}$. 
Using Dynamic Puzzle Matching

Think of: \( k = 4 \) and \(|Q| \approx \sqrt{m}\).

\[ |T_j| = m + \mathcal{O}(k) \]
Think of: \( k = 4 \) and \( |Q| \approx \sqrt{m} \).
Using Dynamic Puzzle Matching

Think of: \( k = 4 \) and \( |Q| \approx \sqrt{m} \).

\[
P \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \ quad
Using Dynamic Puzzle Matching

Think of: \( k = 4 \) and \( |Q| \approx \sqrt{m} \).

Goal: Iterate over all \( I_j \)'s in a DPM instance.

(The leading and trailing pairs are treated separately.)
Using Dynamic Puzzle Matching

Think of: \( k = 4 \) and \(|Q| \approx \sqrt{m}\).
Using Dynamic Puzzle Matching

Think of: $k = 4$ and $|Q| \approx \sqrt{m}$. 

I_{1} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \}
Think of: $k = 4$ and $|Q| \approx \sqrt{m}$. 

Using Dynamic Puzzle Matching
Using Dynamic Puzzle Matching

Think of: $k = 4$ and $|Q| \approx \sqrt{m}$. 
Using Dynamic Puzzle Matching

Think of: $k = 4$ and $|Q| \approx \sqrt{m}$.

We only need to update $\mathcal{O}(k)$ pairs; there has to be a pair $\neq (Q, Q)$ involved!
Using Dynamic Puzzle Matching

Think of: \( k = 4 \) and \( |Q| \approx \sqrt{m} \).

We only need to update \( \mathcal{O}(k) \) pairs; there has to be a pair \( \neq (Q, Q) \) involved!

Over the \( \Theta(\sqrt{m}) \) shifts of \( P \), we need \( \mathcal{O}(\sqrt{m} \cdot k) \) DPM-updates.
Using Dynamic Puzzle Matching

Think of: \( k = 4 \) and \(|Q| \approx \sqrt{m}\).

We only need to update \( \mathcal{O}(k) \) pairs; there has to be a pair \( \neq (Q, Q) \) involved!

Over the \( \Theta(\sqrt{m}) \) shifts of \( P \), we need \( \mathcal{O}(\sqrt{m} \cdot k) \) DPM-updates.

Yields \( \tilde{\mathcal{O}}(k^3 + \sqrt{m} \cdot k^2) \).
$O(k^3)$ DPM-updates via Primitivity
\(O(k^3)\) DPM-updates via Primitivity

\(k = 2\)

\(\mathcal{I}_1\) \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \}

\(\mathcal{I}_2\) \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \}

\cdots

\(\mathcal{I}_{48}\) \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \}

\(\mathcal{I}_{49}\) \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \}

\(\mathcal{I}_{50}\) \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \}

\(\mathcal{I}_{51}\) \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \}

\(\mathcal{I}_{52}\) \{ \{ \} \} \{ \{ \} \} \{ \{ \} \} \{ \{ \} \}
For a plain run \((Q, Q)^y\), at least \(y - k\) copies of \(Q\) will be matched exactly in a \(k\)-error occurrence.
\( \mathcal{O}(k^3) \) DPM-updates via Primitivity

\( k = 2 \)

For a plain run \((Q, Q)^y\), at least \( y - k \) copies of \( Q \) will be matched exactly in a \( k \)-error occurrence.

Cap exponents of plain runs at \( k + 1 \).
\(O(k^3)\) DPM-updates via Primitivity

For a plain run \((Q, Q)^y\), at least \(y - k\) copies of \(Q\) will be matched exactly in a \(k\)-error occurrence.

Cap exponents of plain runs at \(k + 1\).

We do not lose or gain any \(k\)-error occs.
$\mathcal{O}(k^3)$ DPM-updates via Primitivity

$k = 2$

$I_1$  \{ \[\hspace{1cm} 50 \[\hspace{1cm} 50

$I_2$  \[\hspace{1cm} 49 \[\hspace{1cm} 51

\ldots

$I_{48}$  \[\hspace{1cm} 3 \[\hspace{1cm} 97

$I_{49}$  \[\hspace{1cm} 2 \[\hspace{1cm} 98

$I_{50}$  \[\hspace{1cm} 99

$I_{51}$  \[\hspace{1cm} 101

$I_{52}$  \[\hspace{1cm} 102

$I'_1$  \{ \[\hspace{1cm} 3 \[\hspace{1cm} 3

$I'_2$  \[\hspace{1cm} 3 \[\hspace{1cm} 3

\ldots

$I'_{48}$  \[\hspace{1cm} 3 \[\hspace{1cm} 3

$I'_{49}$  \[\hspace{1cm} 2 \[\hspace{1cm} 3

$I'_{50}$  \[\hspace{1cm} 3

$I'_{51}$  \[\hspace{1cm} 3

$I'_{52}$  \[\hspace{1cm} 3
The shown pair of special tiles implies $\mathcal{O}(k)$ DPM-updates.
\( \mathcal{O}(k^3) \) DPM-updates via Primitivity

\( k = 2 \)

\[ \begin{array}{cccccc}
I_1 & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I_2 & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
\vdots & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I_{48} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I_{49} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I_{50} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I_{51} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I_{52} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
\end{array} \]

\[ \begin{array}{cccccc}
I'_1 & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I'_2 & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
\vdots & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I'_{48} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I'_{49} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I'_{50} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I'_{51} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
I'_{52} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} & \{ \square \} \\
\end{array} \]

The shown pair of special tiles implies \( \mathcal{O}(k) \) DPM-updates.

We have \( \mathcal{O}(k^2) \) pairs of special tiles!
$\mathcal{O}(k^3)$ DPM-updates via Primitivity

$k = 2$

$I_1 \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 50$

$I_2 \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 49$

\vdots

$I_{48} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

$I_{49} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 2$

$I_{50} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

$I_{51} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 101$

$I_{52} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 102$

$I'_1 \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

$I'_2 \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

\vdots

$I'_{48} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

$I'_{49} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

$I'_{50} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

$I'_{51} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

$I'_{52} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} \{ \square \} 3$

Alternative $\tilde{\mathcal{O}}(k^4)$-time algorithm!
Overview for $O(k^{2.5})$ DPM-updates

We may get false positives when we have $\geq \sqrt{k}$ edits in a run of $(Q, Q)$. In this case, we must be saving $\geq \sqrt{k}$ by canceling out errors between $P^{\infty}$ and $Q^{\infty}$ with errors between $T^{\infty}$ and $Q^{\infty}$.

We quantify potential savings using a marking scheme based on overlaps of special tiles and verify $O(k^{2.5})$ positions with $\geq \sqrt{k}$ marks using known techniques.

This yields $O(k^{2.5})$ DPM-updates and hence $\tilde{O}(k^{3.5})$ time overall.
Overview for $\mathcal{O}(k^{2.5})$ DPM-updates

Cap exponents of plain runs at $\sqrt{k}$. 
Cap exponents of plain runs at $\sqrt{k}$.

We may get false positives when we have $\geq \sqrt{k}$ edits in a run of $(Q, Q)$. 
Overview for $O(k^{2.5})$ DPM-updates

Cap exponents of plain runs at $\sqrt{k}$.

We may get false positives when we have $\geq \sqrt{k}$ edits in a run of $(Q, Q)$.

\[
\begin{array}{cccccc}
\text{P} & \cdot & \cdot & \cdot & \text{pink} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\\
\text{T} & \cdot & \cdot & \cdot & \text{purple} & \text{yellow} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{array}
\]
Overview for $O(k^{2.5})$ DPM-updates

Cap exponents of plain runs at $\sqrt{k}$.

We may get false positives when we have $\geq \sqrt{k}$ edits in a run of $(Q, Q)$.

![Diagram showing $\sqrt{k}$ exponents in runs $P'$ and $T'$]
Overview for $O(k^{2.5})$ DPM-updates

Cap exponents of plain runs at $\sqrt{k}$.

We may get false positives when we have $\geq \sqrt{k}$ edits in a run of $(Q, Q)$.

\[ \text{Cost: } o + o + \sqrt{k} \cdot \delta_E(Q, \text{rot}^{2k/3}(Q)). \]

P. Charalampopoulos, T. Kociumaka, P. Wellnitz  
Faster Pattern Matching under Edit Distance
Overview for $O(k^{2.5})$ DPM-updates

Cap exponents of plain runs at $\sqrt{k}$.

We may get false positives when we have $\geq \sqrt{k}$ edits in a run of $(Q, Q)$.

In this case, we must be saving $\geq \sqrt{k}$ by canceling out errors between $P$ and $Q^\infty$ with errors between $T$ and $Q^\infty$. 
Overview for $O(k^{2.5})$ DPM-updates

Cap exponents of plain runs at $\sqrt{k}$.

We may get false positives when we have $\geq \sqrt{k}$ edits in a run of $(Q, Q)$.

In this case, we must be saving $\geq \sqrt{k}$ by canceling out errors between $P$ and $Q^\infty$ with errors between $T$ and $Q^\infty$.

We quantify potential savings using a marking scheme based on overlaps of special tiles and verify $O(k^{2.5})$ positions with $\geq \sqrt{k}$ marks using known techniques.
Overview for $\mathcal{O}(k^{2.5})$ DPM-updates

Cap exponents of plain runs at $\sqrt{k}$.

We may get false positives when we have $\geq \sqrt{k}$ edits in a run of $(Q, Q)$.

In this case, we must be saving $\geq \sqrt{k}$ by canceling out errors between $P$ and $Q^\infty$ with errors between $T$ and $Q^\infty$.

We quantify potential savings using a marking scheme based on overlaps of special tiles and verify $\mathcal{O}(k^{2.5})$ positions with $\geq \sqrt{k}$ marks using known techniques.

This yields $\mathcal{O}(k^{2.5})$ DPM-updates and hence $\tilde{O}(k^{3.5})$ time overall.
A Solution to DPM and a Grid View
A Solution to DPM and a Grid View

Faster Pattern Matching under Edit Distance
Theorem [Tiskin; Algorithmica 2015] Matrix $C$ can be computed from (small representations of) $n \times n$ matrices $A$ and $B$ in $O(n \log n)$ time.
A Solution to DPM and a Grid View

\[ P = 10, T_j = 12, k = 2. \]

Only \(|T_j| - |P| + 2k + 1 = \mathcal{O}(k)\) diagonals are relevant.
Preprocessing: Build distance matrices for these small alignment grids.
Preprocessing: Build distance matrices for these small alignment grids.

Update: Maintain a balanced binary tree over them, **stitching** them together.
A Solution to DPM and a Grid View

Preprocessing: Build distance matrices for these small alignment grids.

Update: Maintain a balanced binary tree over them, stitching them together.

Each stitching operation takes $\tilde{O}(k)$ time.
Final Remarks and Open Problems

What is the right exponent?

Cole and Hariharan’s conjecture:

\[ O(n + k \cdot n/m) \]

should be possible.

Is the decision version easier?

What if we allow for some approximation by also reporting an arbitrary subset of the positions in \( \text{Occ}_E(k(P, T)) \) for a small \( \epsilon > 0 \)?

We report starting positions. How fast can we report substrings?
Final Remarks and Open Problems

What is the right exponent?

Cole and Hariharan’s conjecture: $O(n + k^3 \cdot n/m)$ should be possible.
Final Remarks and Open Problems

What is the right exponent?

Cole and Hariharan’s conjecture: $O(n + k^3 \cdot n/m)$ should be possible.

Is the decision version easier?
Final Remarks and Open Problems

What is the right exponent?

Cole and Hariharan’s conjecture: $O(n + k^3 \cdot n/m)$ should be possible.

Is the decision version easier?

What if we allow for some approximation by also reporting an arbitrary subset of the positions in $\mathcal{O}_{(1+\epsilon)k}(P, T) \setminus \mathcal{O}_{k}(P, T)$ for a small $\epsilon > 0$?
Final Remarks and Open Problems

What is the right exponent?

Cole and Hariharan’s conjecture: $\mathcal{O}(n + k^3 \cdot n/m)$ should be possible.

Is the decision version easier?

What if we allow for some approximation by also reporting an arbitrary subset of the positions in $\text{Occ}^E_{(1+\epsilon)k}(P, T) \setminus \text{Occ}^E_k(P, T)$ for a small $\epsilon > 0$?

We report starting positions. How fast can we report substrings?
The End

Thank you for your attention!