Chinese Systematic Naming Saturated Chain Hydrocarbon (SCH) by Graph Algorithms

George [redacted]

[redacted]@bit.edu.cn

MonkeyHouse.Info

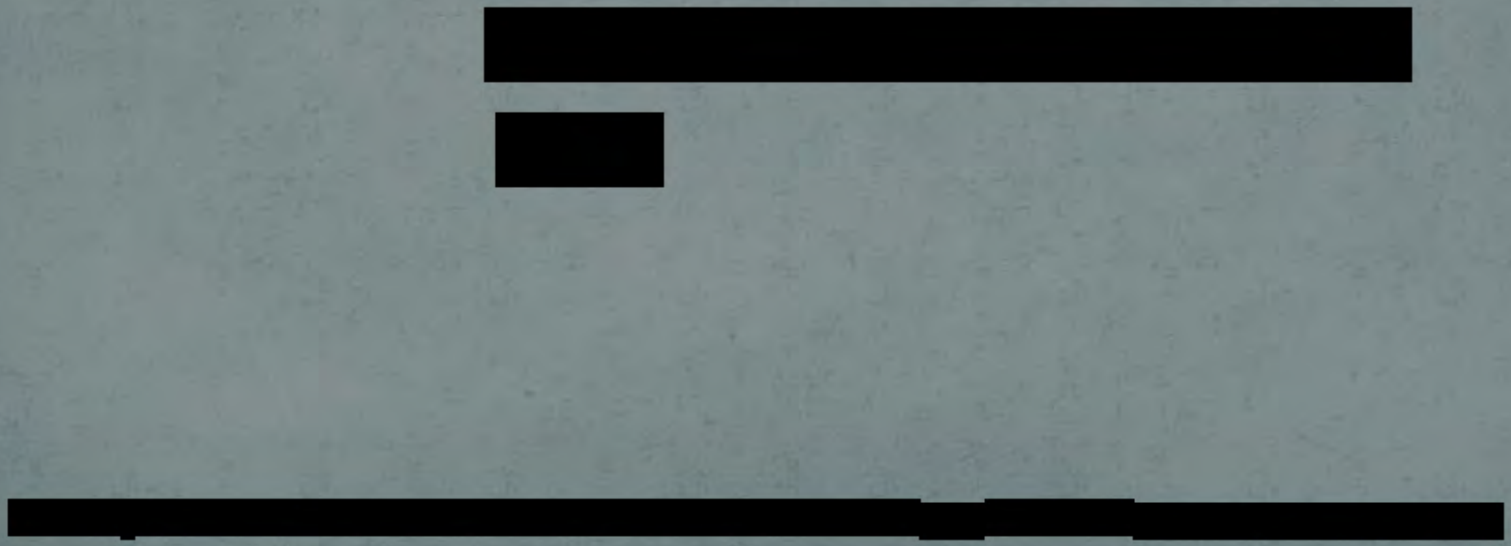
[redacted]

Outline

- Overview
- Transformation
- Algorithms and its Optimizing
- Conclusions and Future Work

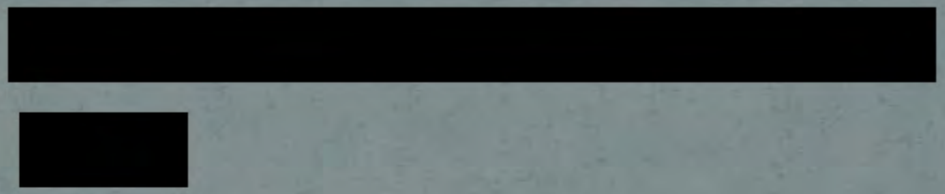
1. Overview

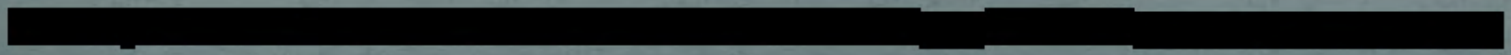
- Problems came from Learning



1. Overview

- Problems came from Learning
- Naming Manually
  - Tedious
  - Low Efficient



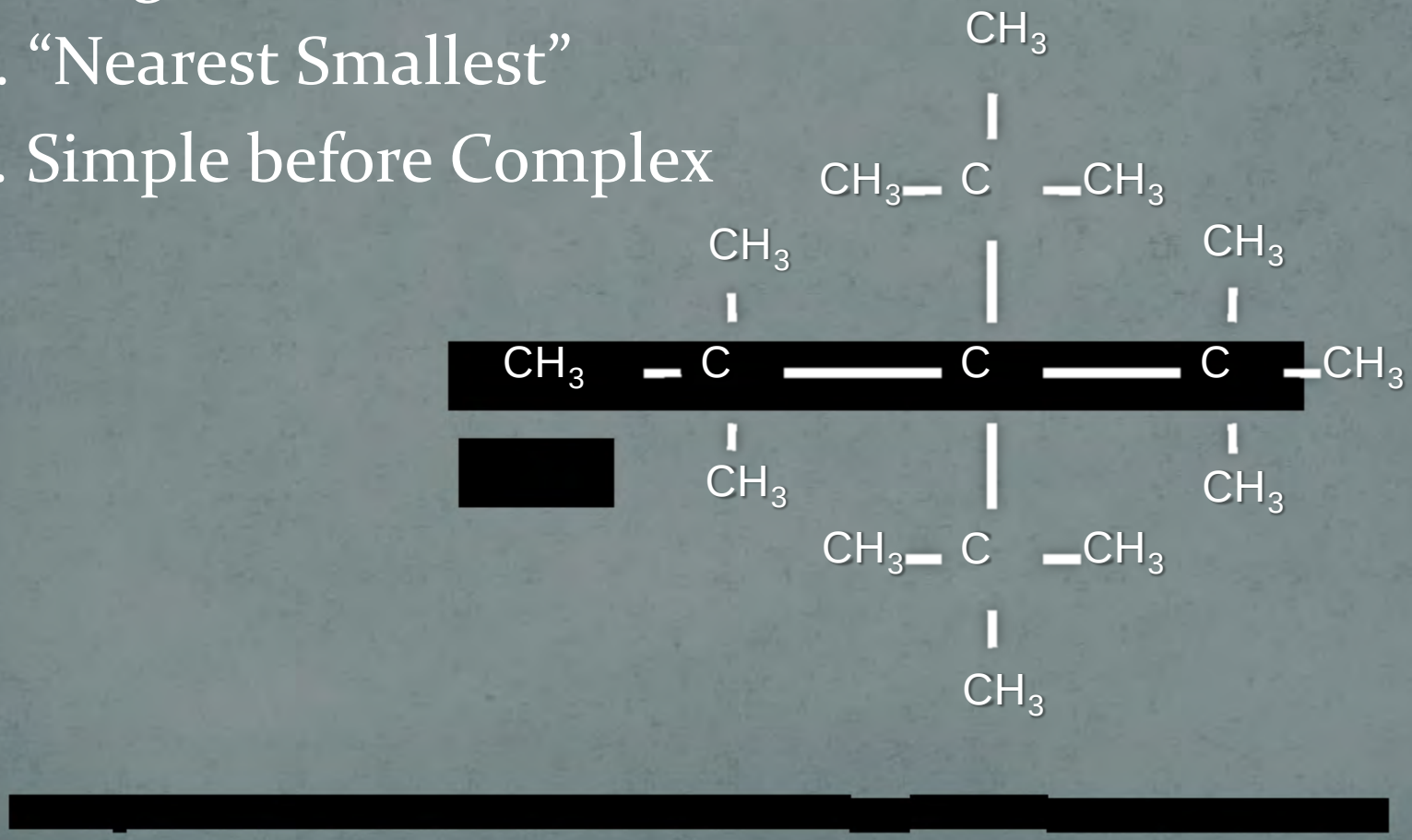


1. Overview

- Problems came from Learning
- Naming Manually
  - Tedious
  - Low Efficient
- Abstracting of Organic Compounds
  - SCH becomes Tree Network
  - Chemistry Problem goes into Graph

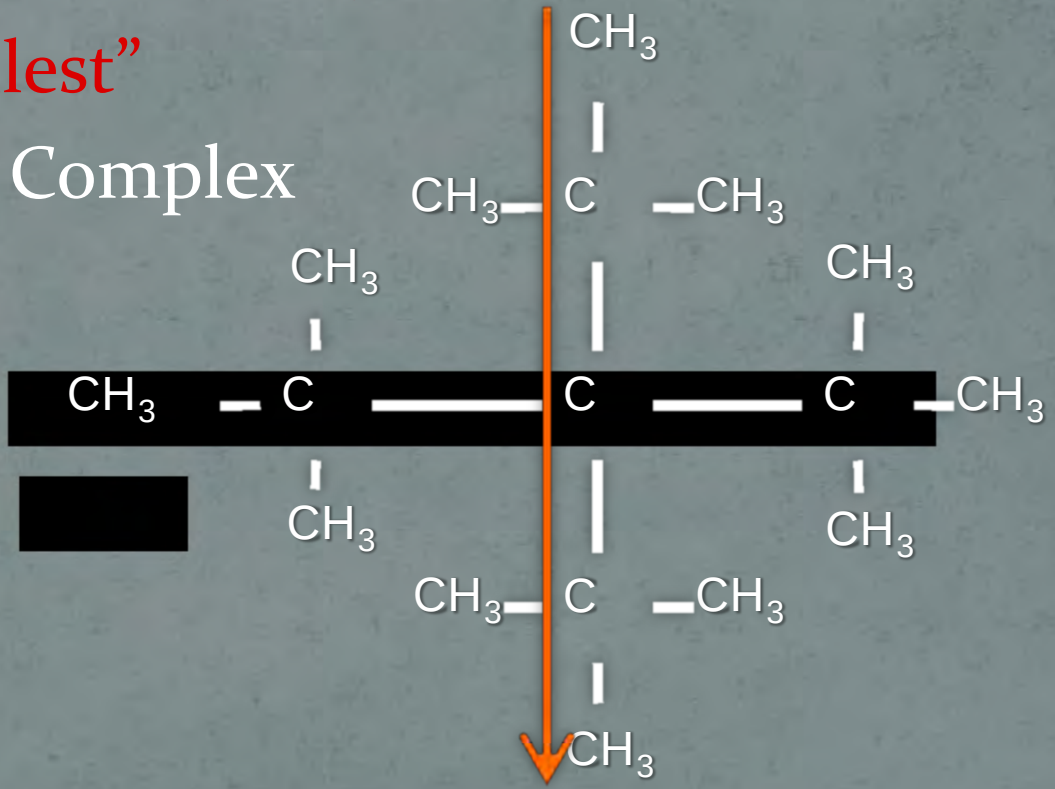
Rules of Chinese Systematic Naming

1. Longest Main Chain (MC)
2. “Nearest Smallest”
3. Simple before Complex



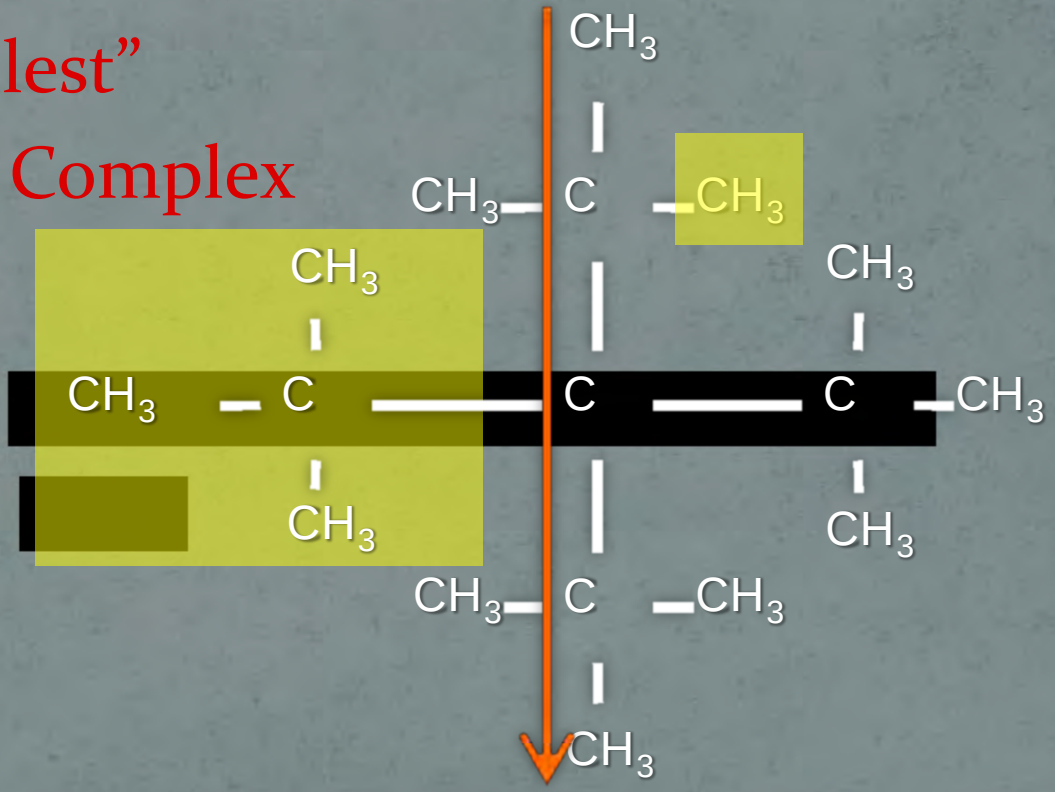
Rules of Chinese Systematic Naming

1. Longest Main Chain (MC)
2. "Nearest Smallest"
3. Simple before Complex



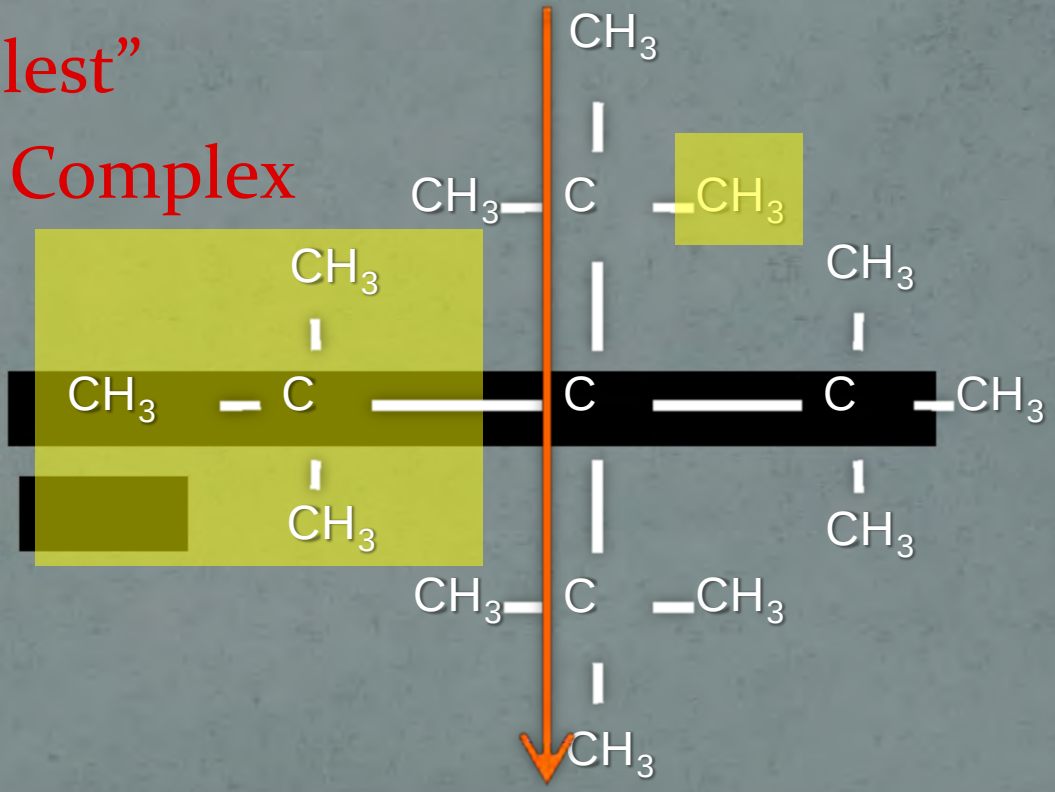
Rules of Chinese Systematic Naming

1. Longest Main Chain (MC)
2. "Nearest Smallest"
3. Simple before Complex



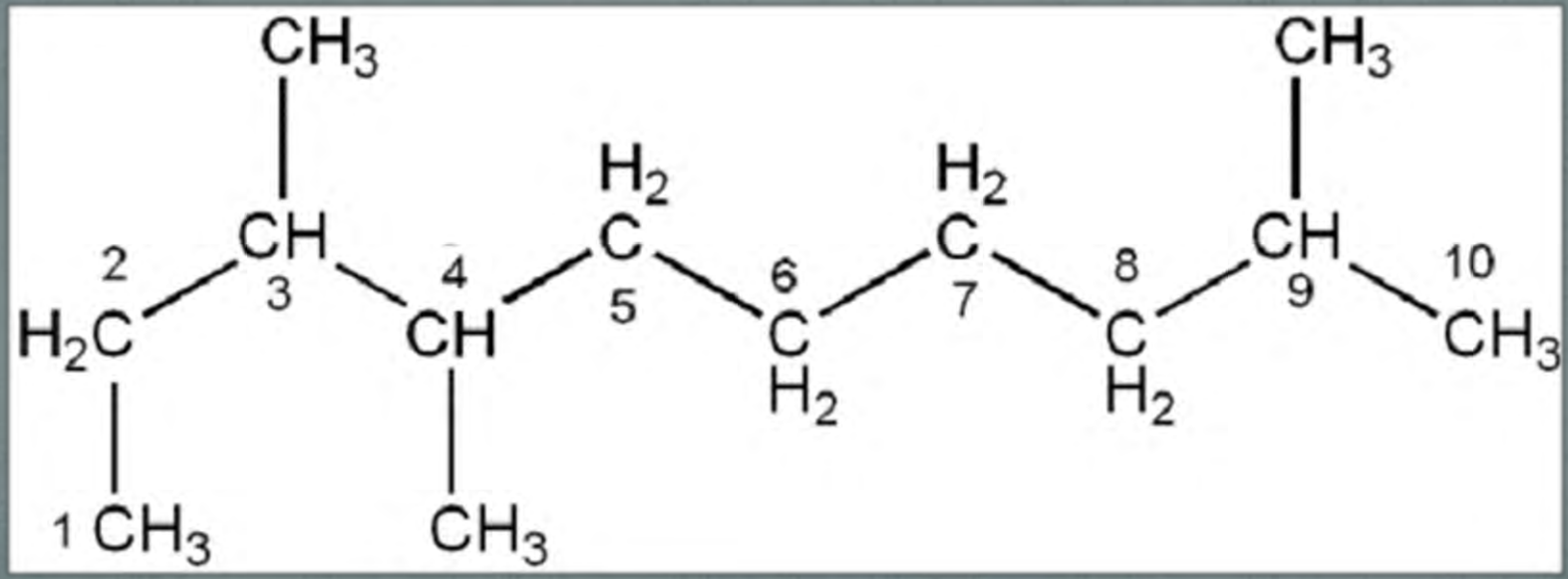
Rules of Chinese Systematic Naming

1. Longest Main Chain (MC)
2. “Nearest Smallest”
3. Simple before Complex



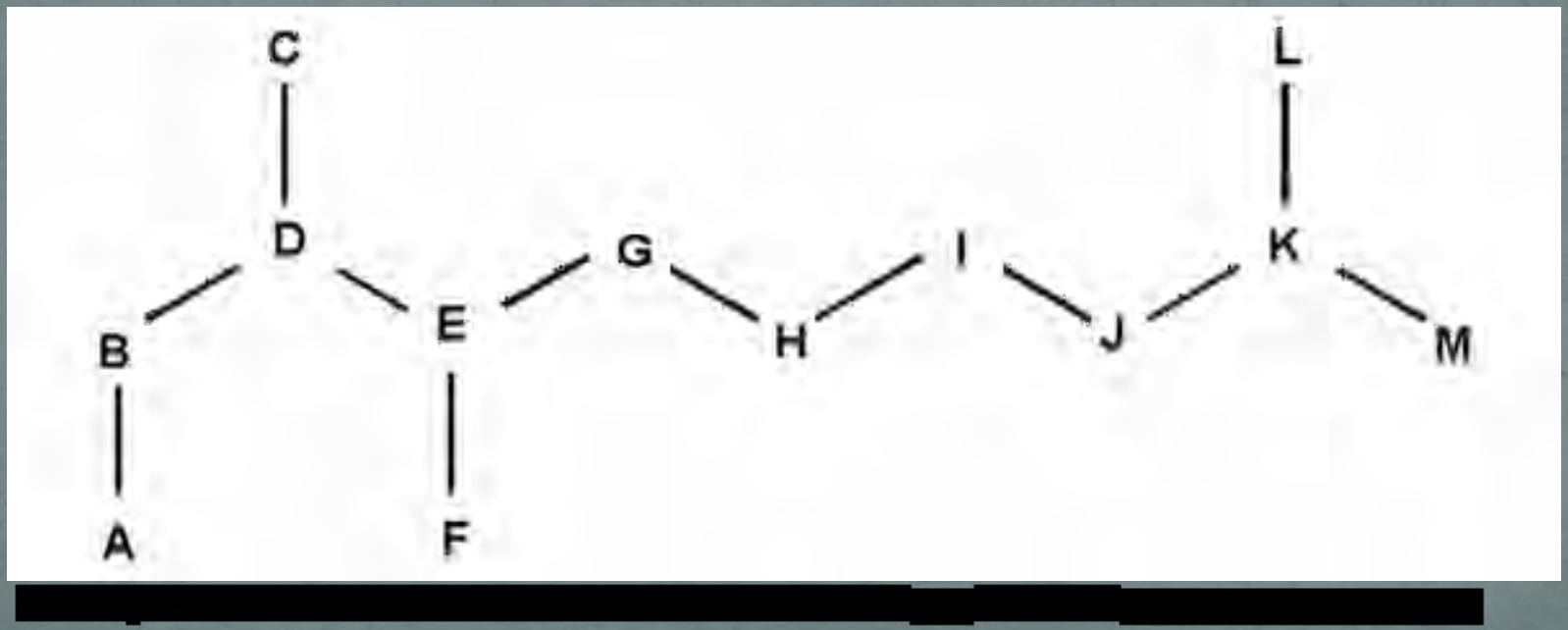
2,2,4,4-四甲基-3,3-二(1,1-二甲基乙基)戊烷

2. Transformation



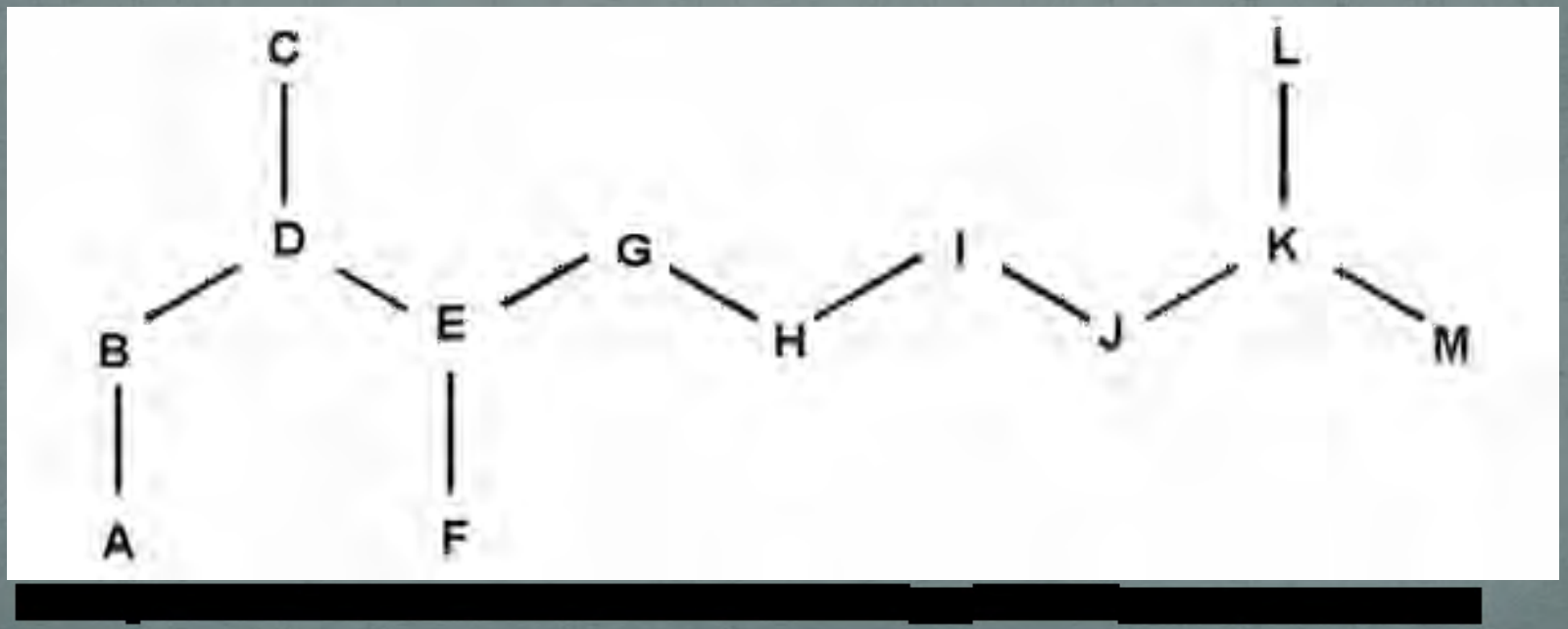
2. Transformation

- Organic Compounds→Graph



2. Transformation

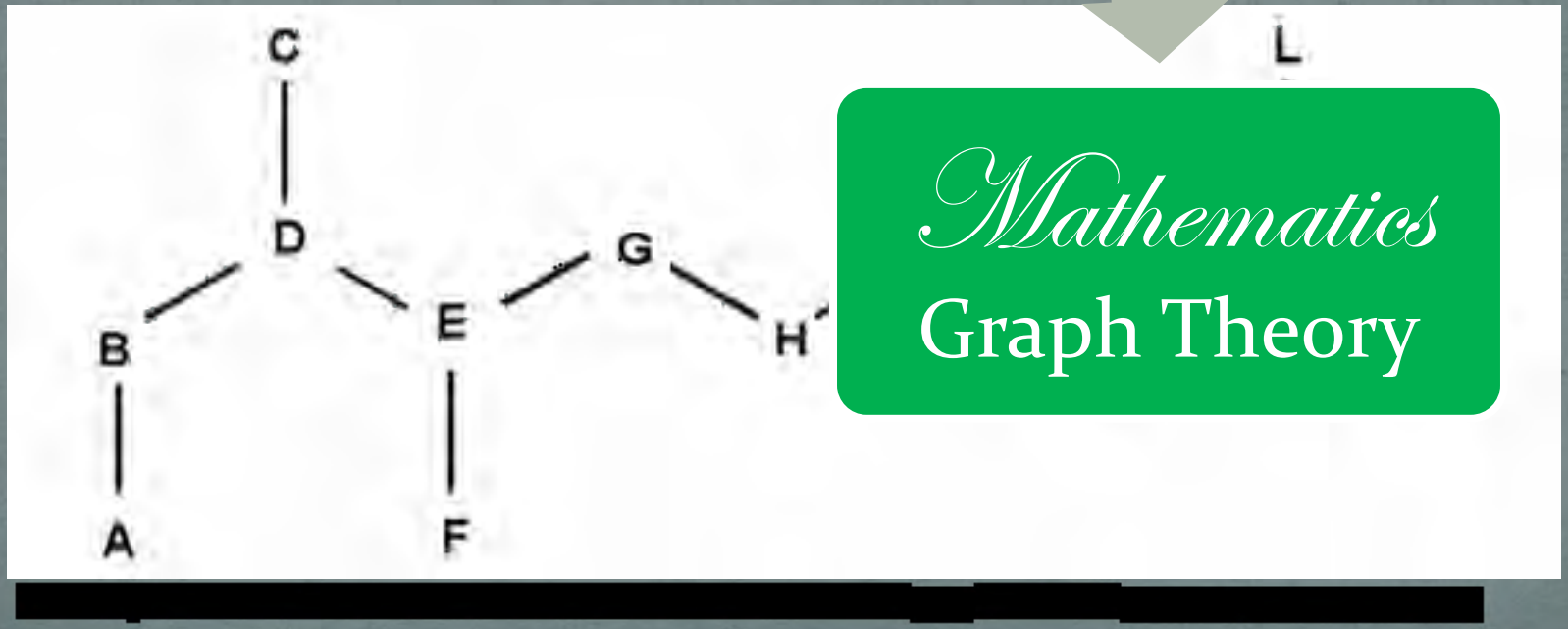
- Organic Compounds→Graph
- SHC→Tree Network



2. Transformation

- Organic Compounds→G
- SHC→Tree Network

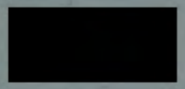
Chemistry
Organic Compounds

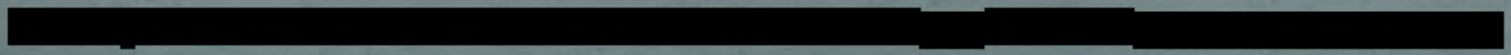


Mathematics
Graph Theory

2. Transformation

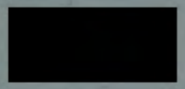
- Longest MC→Diameter of Graph
- Best Naming→Simulation & Comparing

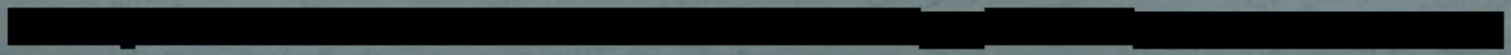




2. Transformation

- Longest MC→Diameter of Graph
- Best Naming→Simulation & Comparing
- Grouping Side Chains→Isomorphic of Graph





2. Transformation

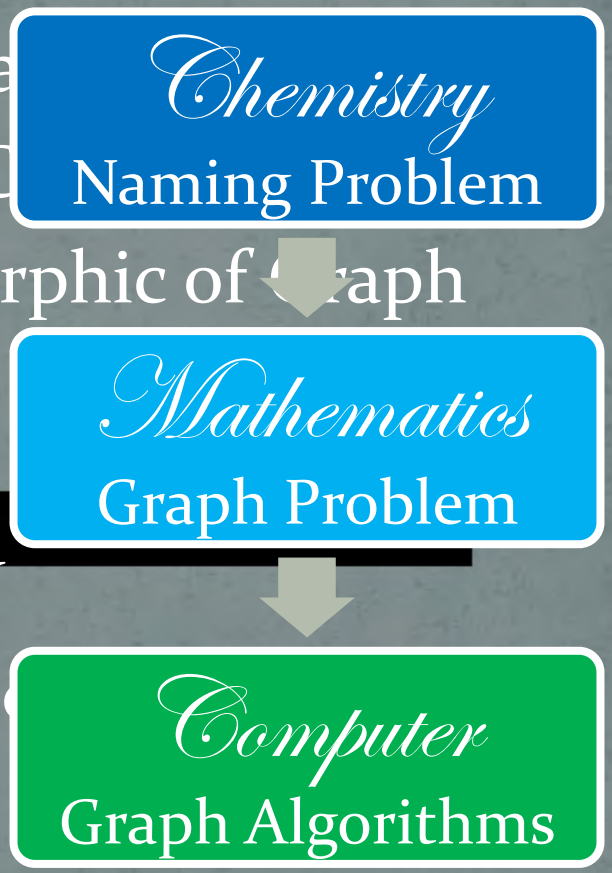
- Longest MC→Diameter of Graph
- Best Naming→Simulation & Comparing
- Grouping Side Chains→Isomorphic of Graph

- SHC→Tree Network (TN)
- Longest MC→Diameter of TN
- Grouping Side Chains (SC)→IsomorphiTrees

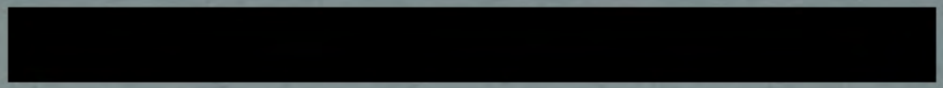
2. Transformation

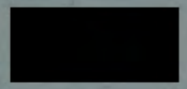
- Longest MC→Diameter of Gra
- Best Naming→Simulation & C
- Grouping Side Chains→Isomorphic of Graph

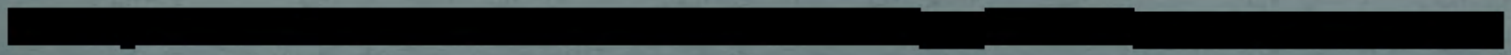
- SHC→Tree Network (TN)
- Longest MC→Diameter of TN
- Grouping Side Chains (SC)→Is



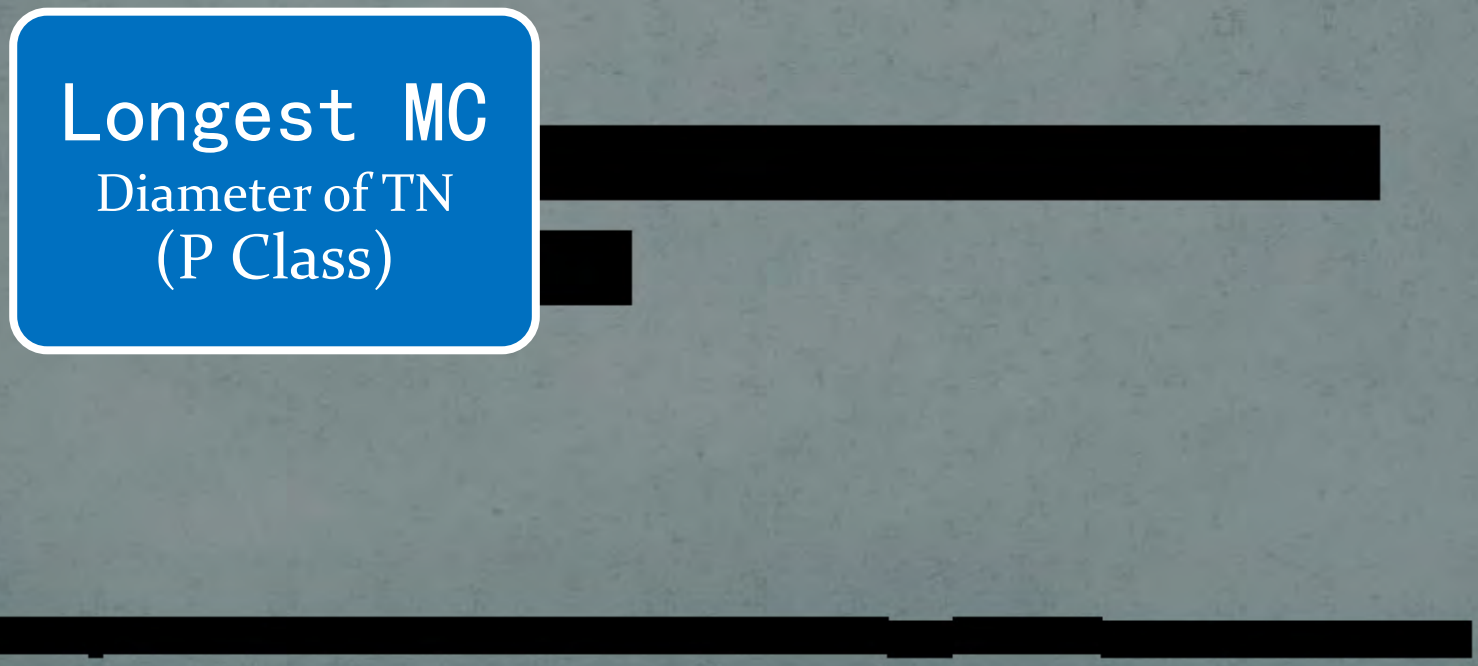
2. Transformation -- Rationality



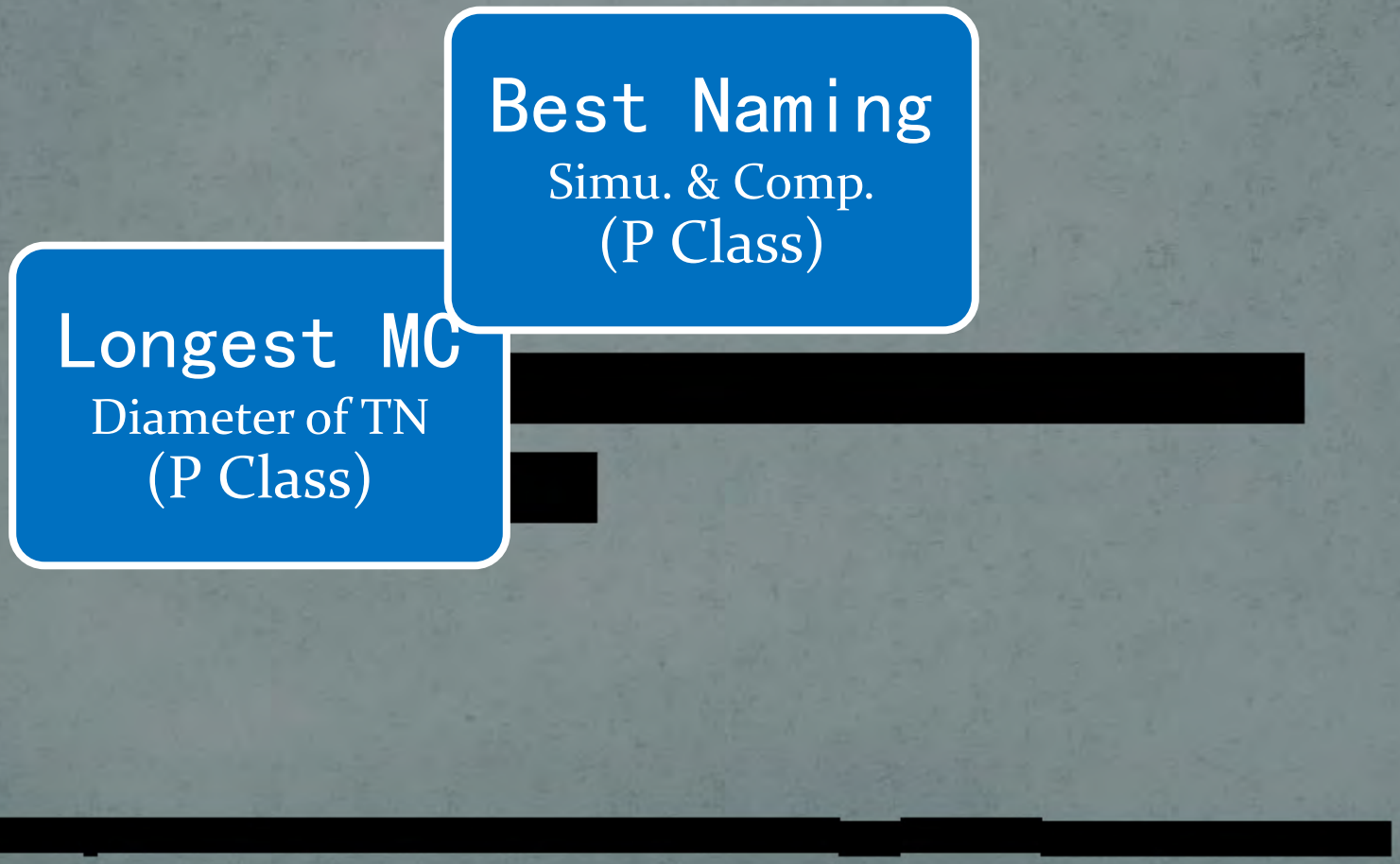




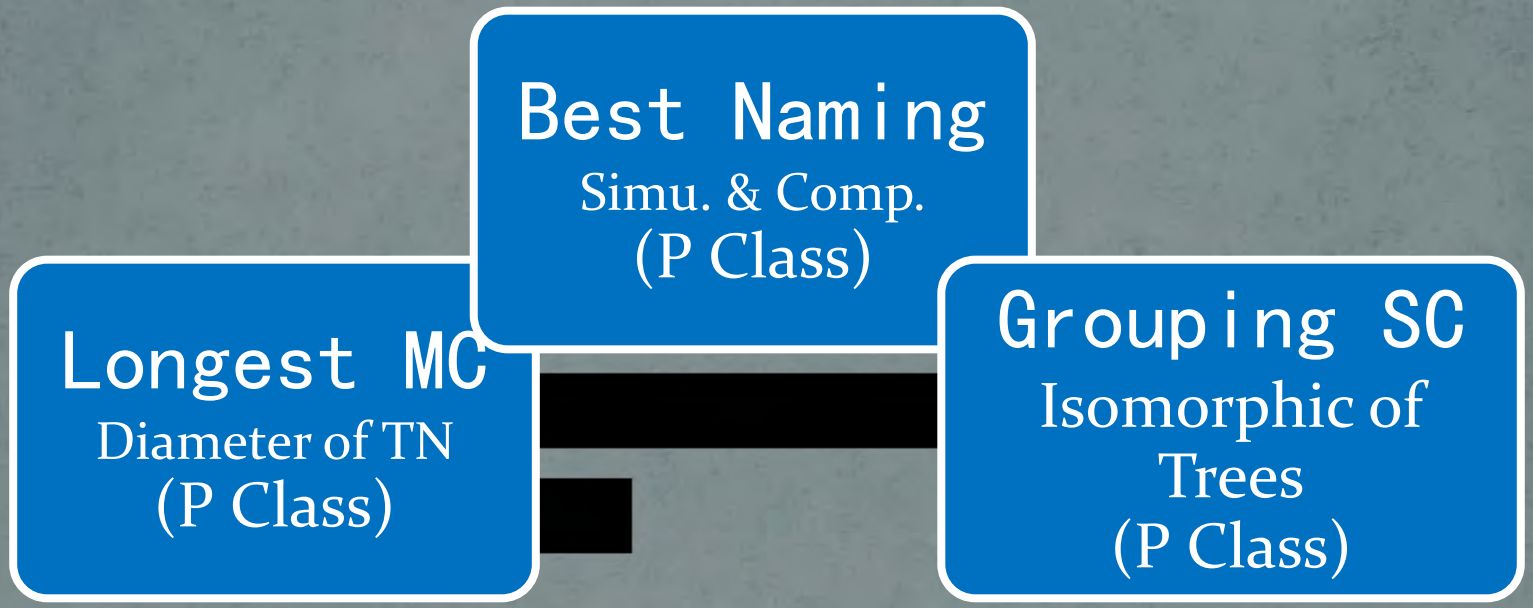
2. Transformation -- Rationality



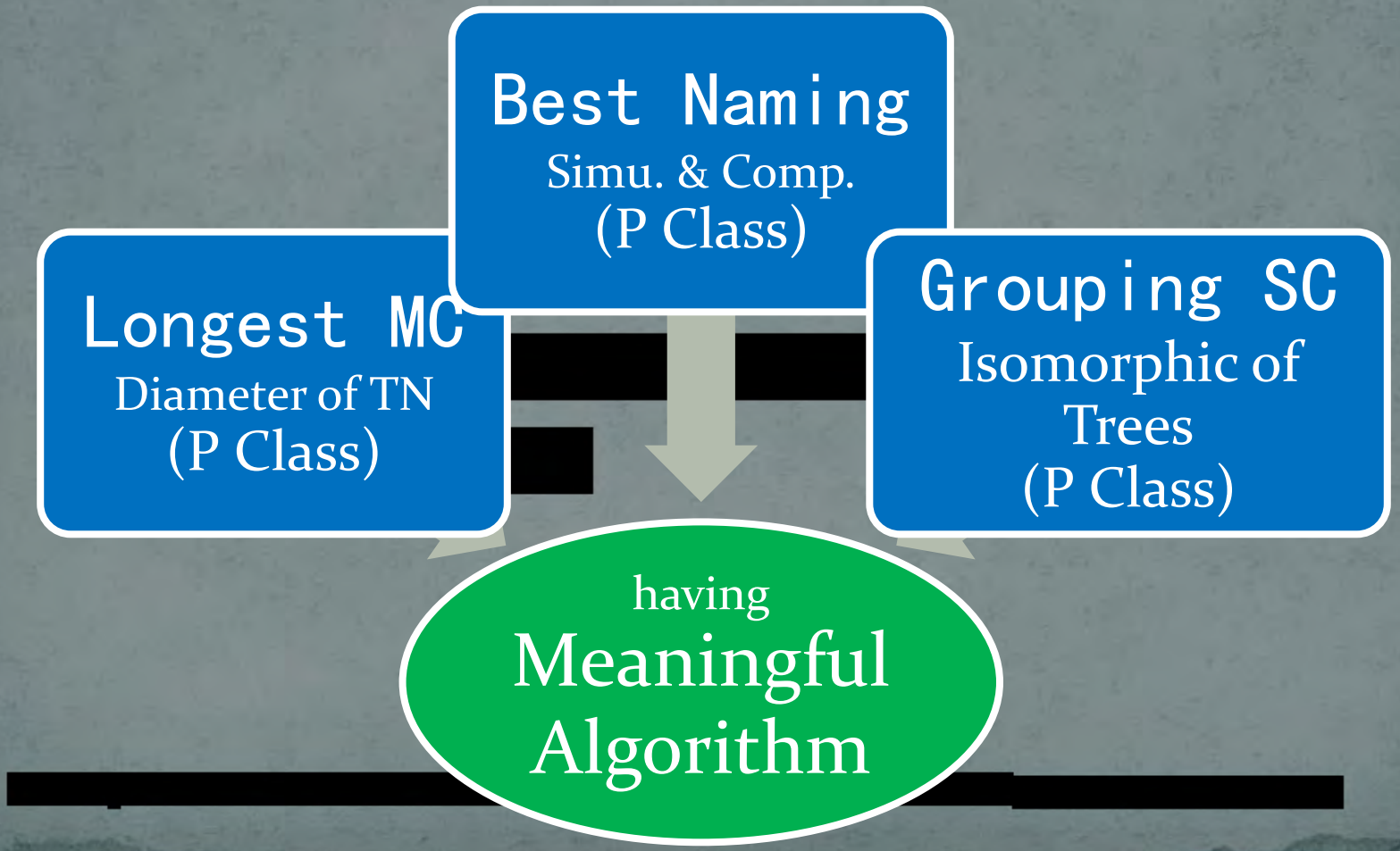
2. Transformation -- Rationality



2. Transformation -- Rationality



2. Transformation -- Rationality

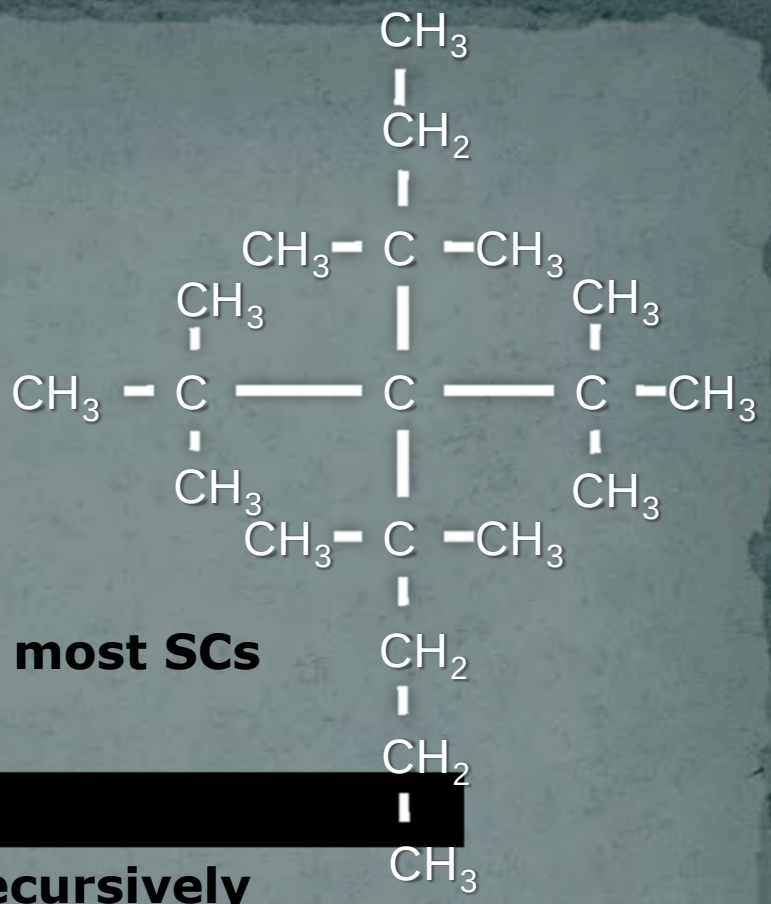


3.1. Algorithms

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$
if 过程$Naming(\mathbb{G})$是第一次调用
 then
  对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理
if $\mathbb{G}$是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将$Count(\mathbb{S})$最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_{\mathbb{Z}}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_{\mathbb{Z}}$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_{\mathbb{S}}$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$
将$\mathbb{M}_{\mathbb{S}_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs**
- **Get Dia. of TN**
- **Get the MC with most SCs**
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs**
- **Choosing the Best Naming**
- **Output in Natural Language**

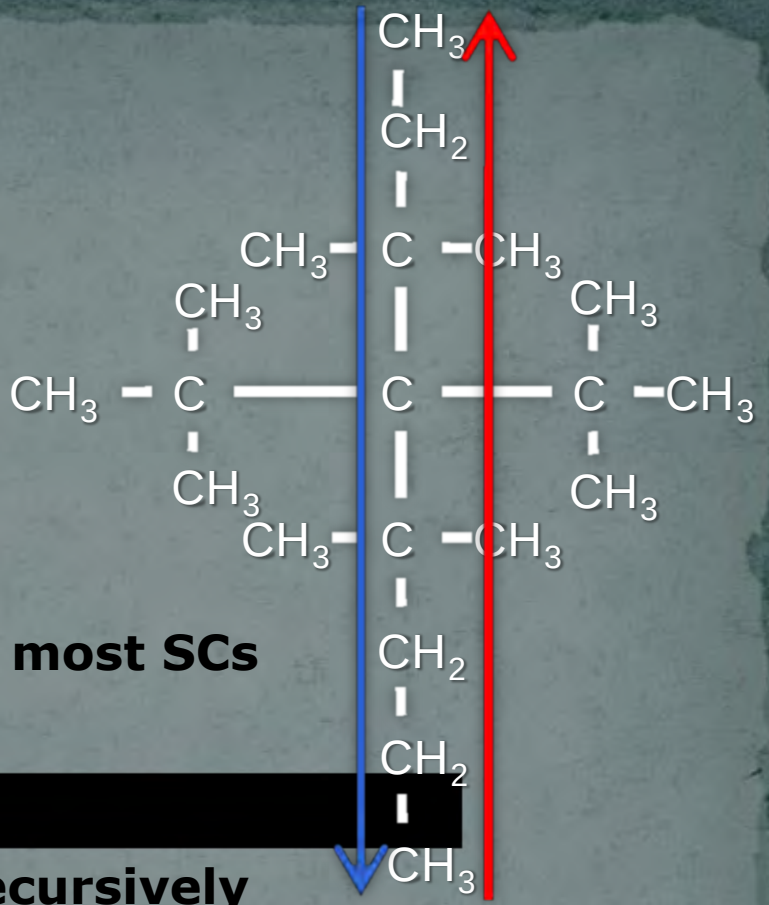


3.1. Algorithms

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$
if 过程$Naming(\mathbb{G})$是第一次调用
 then
  对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理
if $\mathbb{G}$是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将$Count(\mathbb{S})$最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_{\mathbb{Z}}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_{\mathbb{Z}}$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_{\mathbb{S}}$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$
将$\mathbb{M}_{\mathbb{S}_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs**
- **Get Dia. of TN**
- **Get the MC with most SCs**
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs**
- **Choosing the Best Naming**
- **Output in Natural Language**

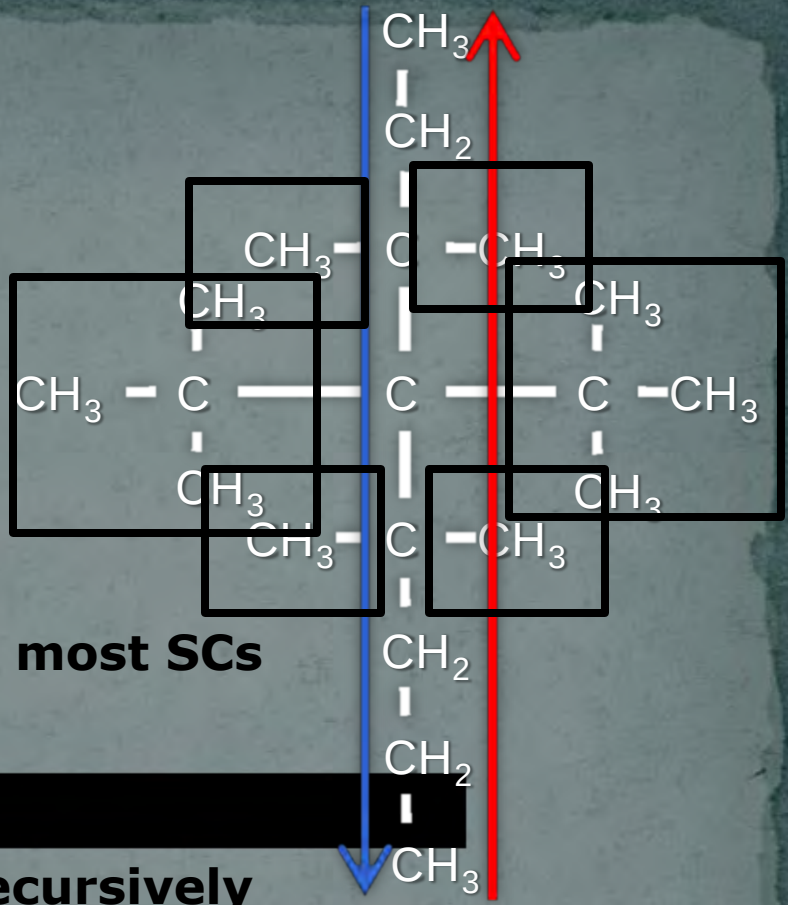


3.1. Algorithms

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷ Procedure $Naming(\mathbb{G})$
if 过程$Naming(\mathbb{G})$是第一次调用
 then
  对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理
if $\mathbb{G}$是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将$Count(\mathbb{S})$最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_{\mathbb{Z}}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_{\mathbb{Z}}$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_{\mathbb{S}}$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$
将$\mathbb{M}_{\mathbb{S}_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs**
- **Get Dia. of TN**
- **Get the MC with most SCs**
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs**
- **Choosing the Best Naming**
- **Output in Natural Language**

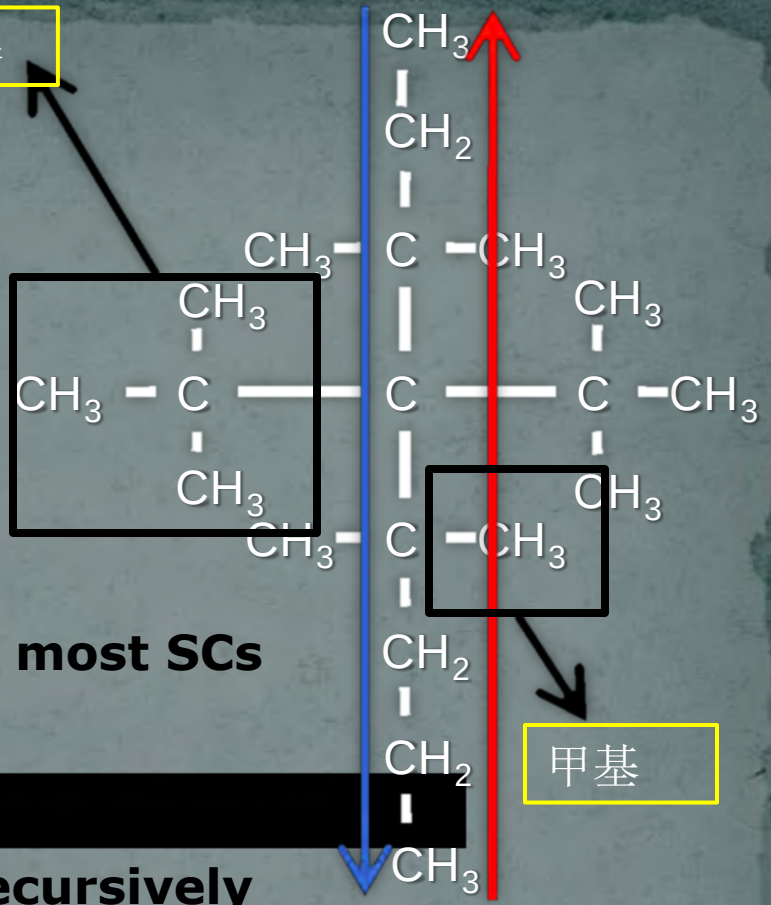


3.1. Algorithms

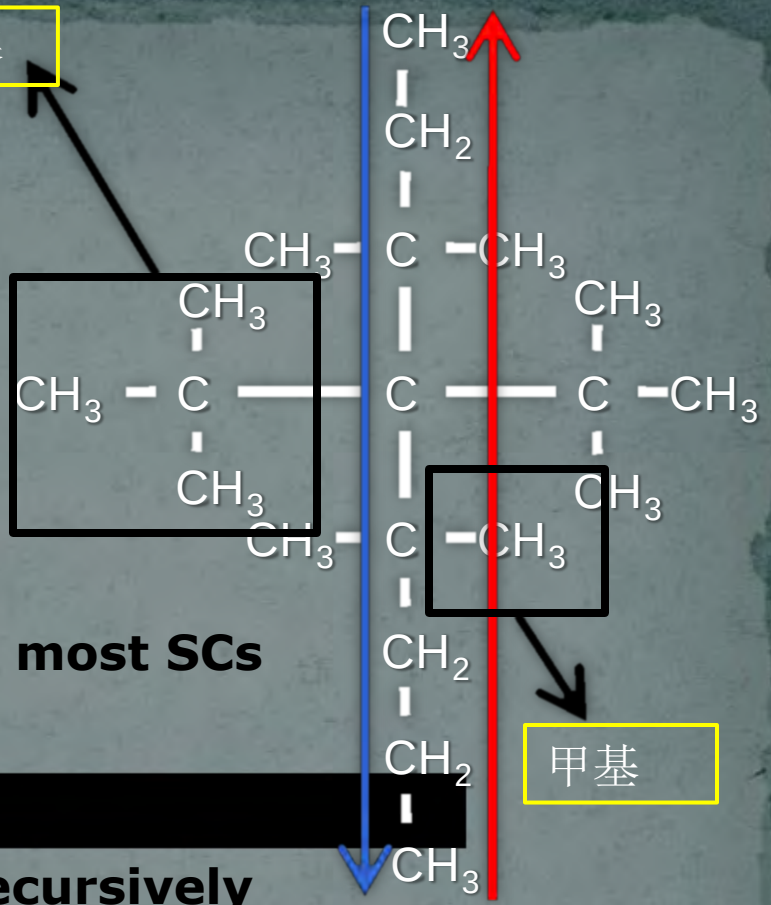
算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure *Naming*(G)
if 过程*Naming*(G)是第一次调用
 then
  对G中所有取代基Z进行复杂性比较的预处理
if G是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将*Count*(S)最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_S$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将Z的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链Z调用*Naming*(Z)进行系统命名$\mathbb{A}_Z$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_Z$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_S$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$S_g$
将M_{S_g}中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs**
- **Get Dia. of TN**
- **Get the MC with most SCs**
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs**
- **Choosing the Best Naming**
- **Output in Natural Language**



3.1. Algorithms



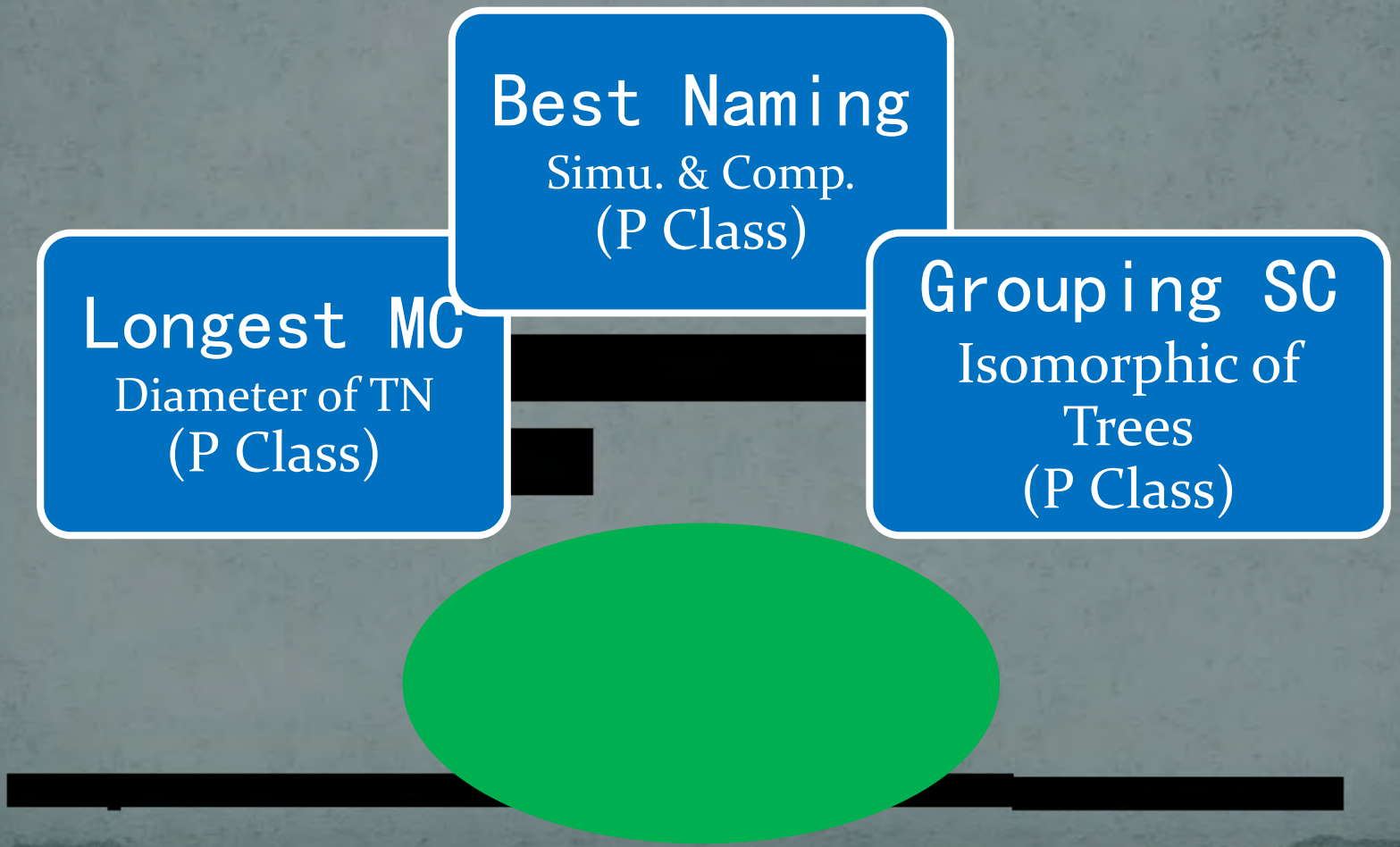
算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$
if 过程$Naming(\mathbb{G})$是第一次调用
 then
  对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理
if $\mathbb{G}$是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将$Count(\mathbb{S})$最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_{\mathbb{Z}}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_{\mathbb{Z}}$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_{\mathbb{S}}$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$
将$\mathbb{M}_{\mathbb{S}_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs**
- **Get Dia. of TN**
- **Get the MC with most SCs**
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs**
- **Choosing the Best Naming**
- **Output in Natural Language**

3,3,5,5 四甲基 4,4 二(1,1-二甲基乙基)
4,4,6,6 四甲基 5,5 二(1,1-二甲基乙基)

2. Transformation -- Rationality

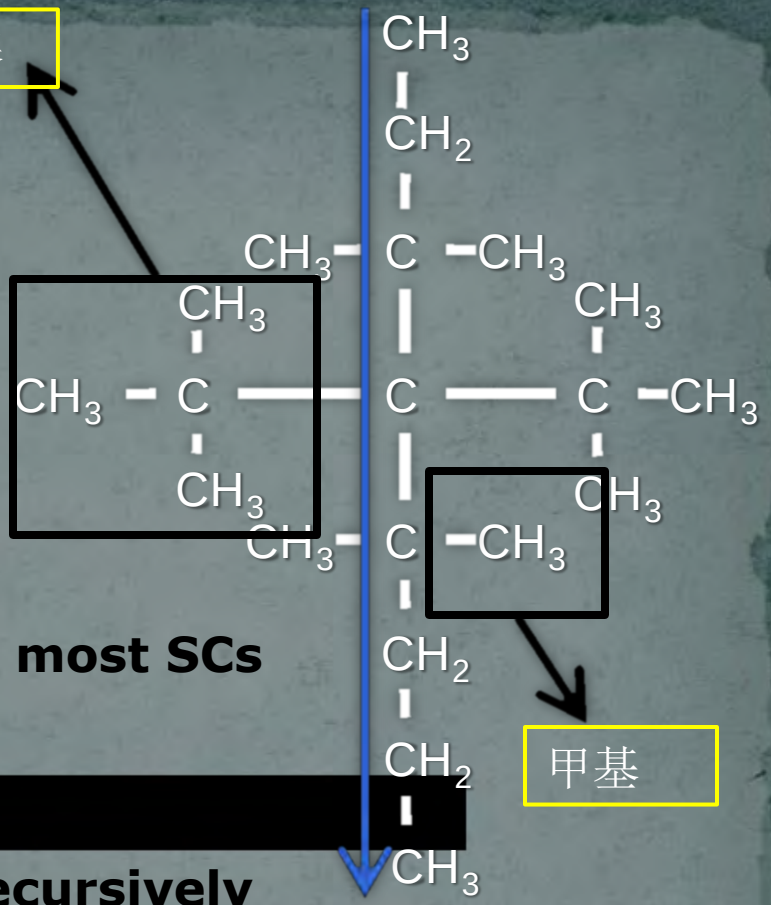


3.1. Algorithms

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$
if 过程$Naming(\mathbb{G})$是第一次调用
 then
  对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理
if $\mathbb{G}$是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将$Count(\mathbb{S})$最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_{\mathbb{Z}}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_{\mathbb{Z}}$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_{\mathbb{S}}$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$
将$\mathbb{M}_{\mathbb{S}_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs**
- **Get Dia. of TN**
- **Get the MC with most SCs**
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs**
- **Choosing the Best Naming**
- **Output in Natural Language**



3,3,5,5-四甲基-4,4-二(1,1-二甲基乙基)辛烷

$$O(m^2) + O(n^2) + O(n^3) + O(m^3) = O(n^3)$$

3.1. Algorithms – Time Complexity

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$
if 过程$Naming(\mathbb{G})$是第一次调用
 then
  对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理
if $\mathbb{G}$是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将$Count(\mathbb{S})$最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_{\mathbb{Z}}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_{\mathbb{Z}}$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_{\mathbb{S}}$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$
将$\mathbb{M}_{\mathbb{S}_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs** $O(m^2)$
- **Get Dia. of TN** $O(n^2)+(2m-1)\times O(n)=O(n^2)$
- **Get the MC with most SCs** $2m \times n \times O(n) = O(n^3)$
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs** $O(m^3)$
- **Choosing the Best Naming**
- **Output in Natural Language**

$$O(m^2) + O(n^2) + O(n^3) + O(m^3) = O(n^3)$$

3.1. Algorithms – Time Complexity

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$
if 过程$Naming(\mathbb{G})$是第一次调用
 then
  对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理
if $\mathbb{G}$是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将$Count(\mathbb{S})$最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_{\mathbb{Z}}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_{\mathbb{Z}}$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_{\mathbb{S}}$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$
将$\mathbb{M}_{\mathbb{S}_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs**
- **Get Dia. of TN**
- **Get the MC with most SCs**
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs**
- **Choosing the Best Naming**
- **Output in Natural Language**

appx. 1k carbons/sec

$O(m^2)$

$O(n^2)+(2m\text{-}1)\times O(n)=O(n^2)$

$2m \times n \times O(n)$
$=O(n^3)$

$O(m^3)$

$$O(m^2) + O(n^2) + O(n^3) + O(m^3) = O(n^3)$$

3.2. Optimized Algorithms

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$
if 过程$Naming(\mathbb{G})$是第一次调用
 then
  对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理
if $\mathbb{G}$是烃
 then
  $\mathbb{D} \leftarrow FindD(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindDJ(\mathbb{G})$
将$Count(\mathbb{S})$最大的直径$\mathbb{S} \in \mathbb{D}$存入集合$\mathbb{R}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_{\mathbb{Z}}$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_{\mathbb{Z}}$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_{\mathbb{S}}$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$
将$\mathbb{M}_{\mathbb{S}_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs** $O(m^2)$
- **Get Dia. of TN** $O(n^2)+(2m-1)\times O(n)=O(n^2)$
- **Get the MC with most SCs** $2m \times n \times O(n) = O(n^3)$
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs** $O(m^3)$
- **Choosing the Best Naming**
- **Output in Natural Language**

$$O(m^2) + O(n^2) + O(n^3) + O(m^3)$$

$$= O(n^3)$$

3.2. Optimized Algorithms

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$

if 过程$Naming(\mathbb{G})$是第一次调用

then

对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理

if $\mathbb{G}$是烃

then

$\mathbb{D} \leftarrow FindD\&Count(\mathbb{G})$

else

$\mathbb{D} \leftarrow FindD\&Count(\mathbb{G})$

将[redacted]最大的直径[redacted]存入集合$\mathbb{R}$

for each $\mathbb{S} \in \mathbb{R}$ do

for each $\mathbb{Z} \in \mathbb{P}_S$ do

if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then

$\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$

将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$

for each $\mathbb{Z} \in \mathbb{K}$ do

对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_Z$

for each $\mathbb{S} \in \mathbb{R}$ do

for each $\mathbb{Z} \in \mathbb{K}$ do

将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_Z$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_S$

通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序

将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$

$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$

将$\mathbb{M}_{S_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs** $O(m^2)$
- **Get MC with longest dia. and most SCs on the same time** $O(n^2)$
- **Grouping SCs** [redacted]
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs** $O(m^2 \log m)$
- **Choosing the Best Naming**
- **Output in Natural Language** [redacted]

$$O(m^2) + O(n^2) + O(n^2) + O(m^2 \log m) = O(n^2 \log n)$$

3.2. Optimized Algorithms

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure $Naming(\mathbb{G})$

if 过程$Naming(\mathbb{G})$是第一次调用

then

对$\mathbb{G}$中所有取代基$\mathbb{Z}$进行复杂性比较的预处理

if $\mathbb{G}$是烃

then

$\mathbb{D} \leftarrow FindD\&Count(\mathbb{G})$

else

$\mathbb{D} \leftarrow FindD\&Count(\mathbb{G})$

将[redacted]最大的直径[redacted]存入集合$\mathbb{R}$

for each $\mathbb{S} \in \mathbb{R}$ do

for each $\mathbb{Z} \in \mathbb{P}_S$ do

if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then

$\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$

将$\mathbb{Z}$的地址存入$Hash(Rank(\mathbb{Z}))$

for each $\mathbb{Z} \in \mathbb{K}$ do

对侧链$\mathbb{Z}$调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_Z$

for each $\mathbb{S} \in \mathbb{R}$ do

for each $\mathbb{Z} \in \mathbb{K}$ do

将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_Z$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_S$

通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序

将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$

$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$\mathbb{S}_g$

将$\mathbb{M}_{S_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs** $O(m^2)$
- **Get MC with longest dia. and most SCs on the same time** $O(n^2)$
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs** $O(m^2 \log m)$
- **Choosing the Best Naming**
- **Output in Natural Language**

$$O(m^2) + O(n^2) + O(n^2) + O(m^2 \log m) = O(n^2 \log n)$$

3.2. Optimized Algorithms

算法 3 饱和链烃和一价饱和链烃基系统命名算法

▷Procedure *Naming*(𝔾)
if 过程*Naming*(𝔾)是第一次调用
 then
  对𝔾中所有取代基ℤ进行复杂性比较的预处理
if 𝔾是烃
 then
  $\mathbb{D} \leftarrow FindD\&Count(\mathbb{G})$
 else
  $\mathbb{D} \leftarrow FindD\&Count(\mathbb{G})$
将[redacted]最大的直径[redacted]存入集合ℝ
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{P}_S$ do
  if $Hash(Rank(\mathbb{Z})) = \text{NULL}$ then
   $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$
  将ℤ的地址存入$Hash(Rank(\mathbb{Z}))$
for each $\mathbb{Z} \in \mathbb{K}$ do
 对侧链ℤ调用$Naming(\mathbb{Z})$进行系统命名$\mathbb{A}_Z$
for each $\mathbb{S} \in \mathbb{R}$ do
 for each $\mathbb{Z} \in \mathbb{K}$ do
  将$Hash(Rank(\mathbb{Z})) \cap \mathbb{S}$的标号与$\mathbb{A}_Z$连接后，存入该直径的侧链系统命名集合$\mathbb{M}_S$
通过$Compare(\mathbb{Z}_1, \mathbb{Z}_2)$对侧链$\mathbb{Z} \in \mathbb{K}$的复杂性进行堆排序
将命名开始编号一端距离最简单侧链最近的直径$\mathbb{S} \in \mathbb{R}$存入集合$\mathbb{R}'$
$\mathbb{S} \in \mathbb{R}'$中，侧链序号和最小的命名为$S_g$
将$\mathbb{M}_{S_g}$中对侧链的命名按照顺序输出并用短线连接，最后输出碳原子数为$card(\mathbb{D})$的直链烷烃的习惯命名

- **Sorting SCs**
- **Get MC with longest dia. and most SCs on the same time**
- **Grouping SCs**
- **Naming SC Recursively**
- **Connecting Numbers and SCs**
- **Sorting SCs** $O(m^2 \log m)$
- **Choosing the Best Naming**
- **Output in Natural Language**

appx. 10k carbons/sec

$O(n^2)$

4.1. Conclusions

- **SHC→Tree Network in Graph Theory**
- **Getting MC and Grouping SCs→Diameters and** isomorphism of TN
- **Give Chinese Systematic Naming Algorithm (CSNA) based on Diameter and Isomorphism Algorithms**
- **Optimize the Algorithm to Get Better Performance**
- **CSNA: Efficient and Fast**

4.2. Future Work

- Accurate Computing Complexity Estimation
- Pattern Matching :
  - Generating TNs from Expressions
- Generalization

Thanks

Thanks for Your Attention [redacted]

[1] [redacted]