

TECHNISCHE UNIVERSITÄT DRESDEN

CENTER FOR INFORMATION SERVICES  
& HIGH PERFORMANCE COMPUTING  
PROF. DR. WOLFGANG E. NAGEL

## Performance Tuning & Parallelisation of Inchworm

Ankur Sharma

Dresden, March 18, 2014



## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Background . . . . .	4
1.2	Objectives . . . . .	4
<b>2</b>	<b>Biological Background</b>	<b>4</b>
2.1	Sequence Assembly . . . . .	4
2.2	Algorithm . . . . .	4
<b>3</b>	<b>Trinity</b>	<b>4</b>
3.1	Structure . . . . .	4
3.2	Phases . . . . .	4
3.2.1	0 - Jellyfish . . . . .	4
3.2.2	1 - Inchworm . . . . .	4
3.2.3	2 - Chrysalis . . . . .	4
3.2.4	3 - Butterfly . . . . .	4
3.3	Applications . . . . .	4
<b>4</b>	<b>Inchworm</b>	<b>4</b>
4.1	Phases . . . . .	4
4.1.1	Parsing . . . . .	4
4.1.2	Sorting . . . . .	4
4.1.3	Pruning . . . . .	4
4.1.4	Assembling . . . . .	4
4.2	Implementation . . . . .	4
4.3	Performance Bottlenecks . . . . .	4
4.4	Conclusion . . . . .	4
<b>5</b>	<b>Performance Optimisation</b>	<b>4</b>
5.1	Parsing . . . . .	4
5.1.1	Memory Mapped IO . . . . .	4
5.1.2	Benefits . . . . .	4
5.1.3	How it is used? . . . . .	4
5.1.4	Impact on Inchworm . . . . .	4
5.2	Sorting . . . . .	4
5.3	Pruning . . . . .	4
5.3.1	Parallel Pruning . . . . .	4
5.3.2	Performance Gain . . . . .	4
5.4	Assembly . . . . .	4
5.4.1	Bottlenecks . . . . .	4
5.4.2	Parallel Assembling . . . . .	4
5.4.3	Benefits . . . . .	4
5.4.4	Impact on inchworm results . . . . .	4
<b>6</b>	<b>Conclusion</b>	<b>4</b>
6.1	Future work . . . . .	4





## **1 Introduction**

### **1.1 Background**

### **1.2 Objectives**

## **2 Biological Background**

### **2.1 Sequence Assembly**

### **2.2 Algorithm**

## **3 Trinity**

### **3.1 Structure**

### **3.2 Phases**

#### **3.2.1 0 - Jellyfish**

#### **3.2.2 1 - Inchworm**

#### **3.2.3 2 - Chrysalis**

#### **3.2.4 3 - Butterfly**

### **3.3 Applications**

## **4 Inchworm**

### **4.1 Phases**

#### **4.1.1 Parsing**

#### **4.1.2 Sorting**

#### **4.1.3 Pruning**

#### **4.1.4 Assembling**

### **4.2 Implementation**

### **4.3 Performance Bottlenecks**

### **4.4 Conclusion**

## **5 Performance Optimisation**

### **5.1 Parsing**

#### **5.1.1 Memory Mapped IO**

#### **5.1.2 Benefits**

#### **5.1.3 How it is used?**

#### **5.1.4 Impact on Inchworm**

### **5.2 Sorting**

### **5.3 Pruning**

#### **5.3.1 Parallel Pruning**

#### **5.3.2 Performance Gain**

### **5.4 Assembly**

#### **5.4.1 Bottlenecks**

#### **5.4.2 Parallelisation**