



中国科学技术大学
University of Science and Technology of China

引 论

《编译原理和技术(H)》、《编译原理(H)》

张昱

0551-63603804, yuzhang@ustc.edu.cn

中国科学技术大学
计算机科学与技术学院



ACM图灵奖

<https://amturing.acm.org/bysubject.cfm>

□ 编程语言、编译相关的获奖者是最多的 占约1/3

Analysis of Algorithms
Combinatorial Algorithms Compilers
Computer Architecture Computer Hardware
Data Structures Databases Education Error Correcting Codes Finite Automata Graphics
Interactive Computing Internet Communications List Processing Numerical Analysis
Numerical Methods Object Oriented Programming Operating Systems Personal Computing
Program Verification Programming
Programming Languages
Verification of Hardware and Software Models Computer Systems Machine Learning
Parallel Computation

Artificial Intelligence

Computational Complexity

Cryptography

Proof Construction Software
Theory Software Engineering





程序语言与编译系统发展的契机

□ 人工智能的再次兴起，2021：人工智能的普及之年

- 人工智能加速芯片
- 人工智能算法开发

} 对程序语言与编译
提出更高要求

□ 国产芯片五年计划，2020年8月

- 到2025年将实现70%的芯片自给率
- 2020年新增超过6万家芯片相关企业



➔ 面向应用/硬件的领域特定语言、软硬件协同的编译系统优化



主要内容

1

编程语言及设计

2

编译器的阶段

3

编译器的作用及形式

4

编译技术的应用与挑战



主要内容

1

编程语言及设计

2

编译器的阶段

3

编译器的作用及形式

4

编译技术的应用与挑战



□ 什么是编程语言

- **A programming language** is a notation for describing computations to people and to machines.

□ 每种编程语言有自己的计算模型

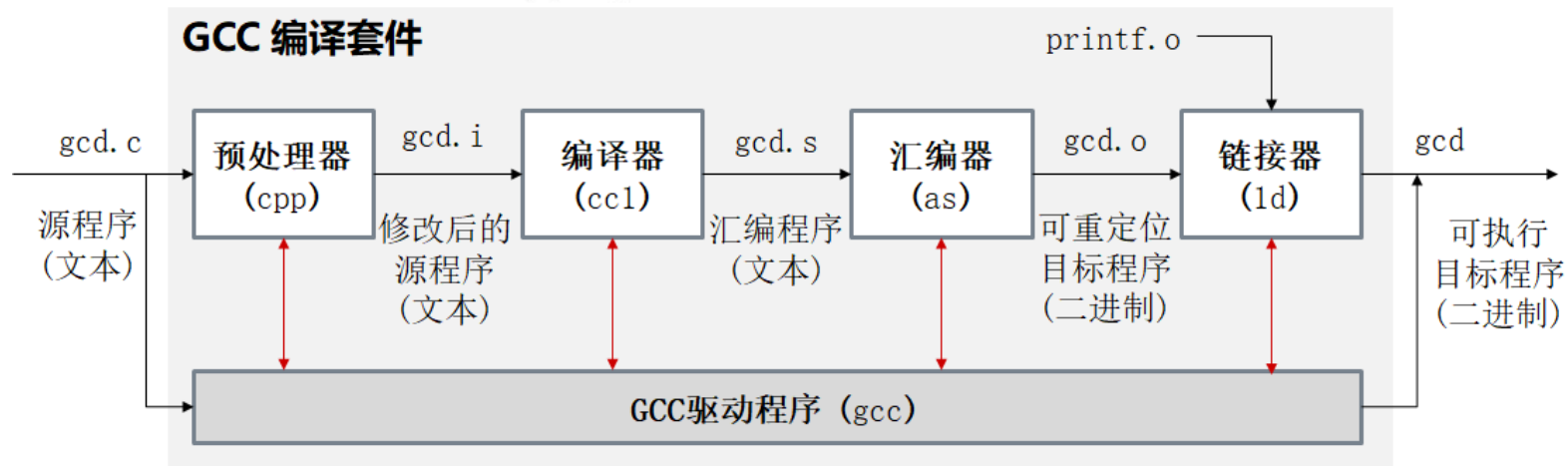
- 过程型(Procedural): **C, C++, C#, Java, Go**
- 声明型(Declarative): **SQL, ...**
- 逻辑型(Logic): **Prolog, ...**
- 函数式(Functional): **Lisp/Scheme, Haskell, ML, OCaml...**
- 脚本型(Scripting): **AWK, Perl, Python, PHP, Ruby, ...**



求最大公约数 gcd

```
int gcd(int a, int b) {  
    while (a != b) {  
        if (a > b) a = a - b;  
        else b = b - a;  
    }  
    return a;  
}
```

// C



```
let rec gcd a b =  
    if a = b then a  
    else if a > b then gcd b (a - b)  
    else gcd a (b - a)
```

(* OCaml *)

```
gcd(A,B,G) :- A = B, G = A.  
gcd(A,B,G) :- A > B, C is A-B, gcd(C,B,G).  
gcd(A,B,G) :- B > A, C is B-A, gcd(C,A,G).
```

% Prolog



□ <https://godbolt.org/>

The screenshot displays the Godbolt online compiler interface with four main panels:

- A. 源代码 (Source Code):** Shows Python source code for a GCD function and its usage. A red arrow points to the language dropdown menu, labeled "选择编程语言" (Select programming language).
- B. 编译: 汇编码 (Compile: Assembly):** Shows the compiled assembly code for Python 3.12. A red arrow points to the compiler dropdown menu, labeled "选择编译器" (Select compiler).
- C. 运行输出 (Execution Output):** Shows the output of the program execution, displaying "gcd(21,36)=3".
- D. LLVM IR输出 (LLVM IR Output):** Shows the LLVM IR output, which is currently empty.

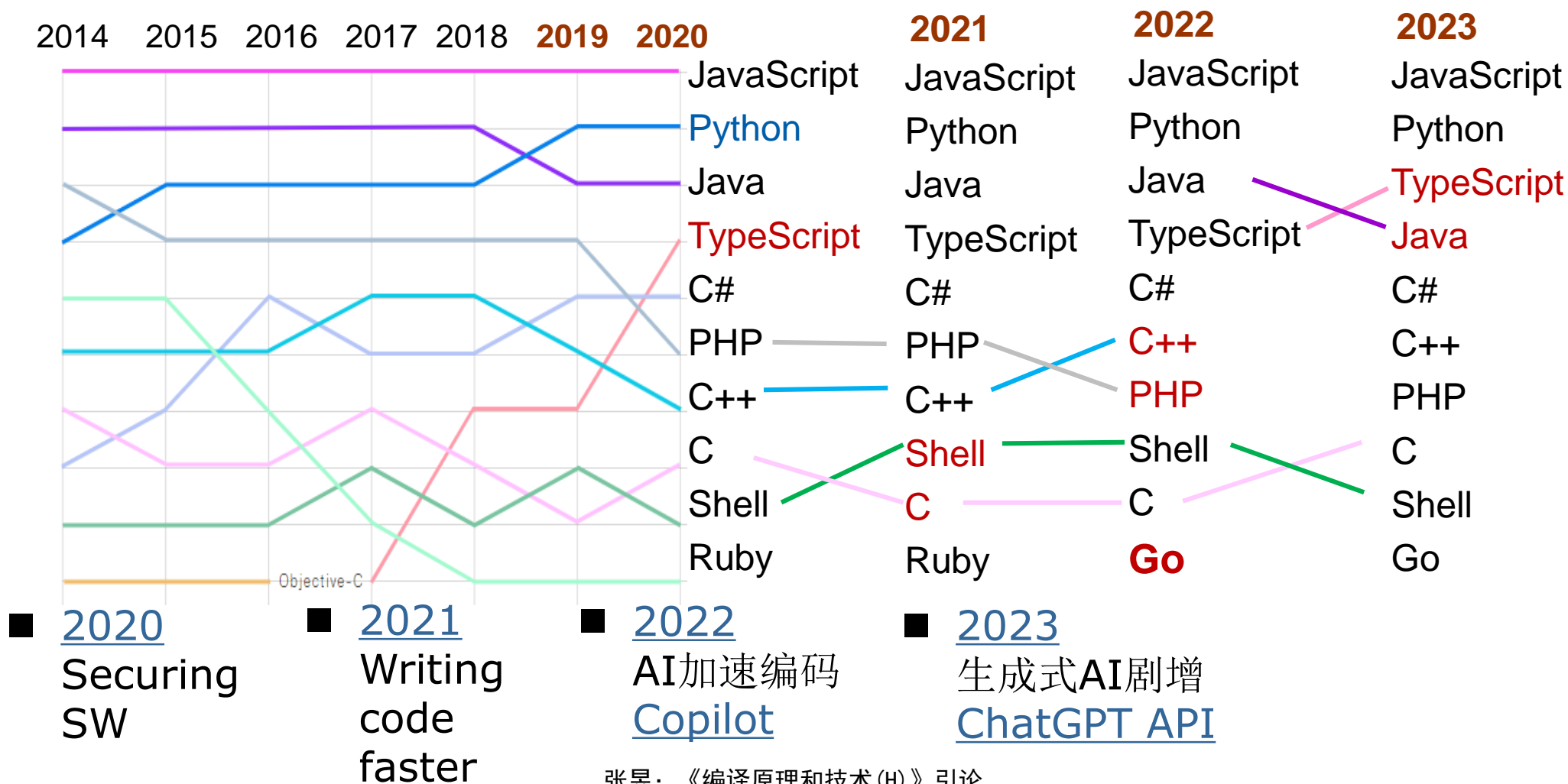
□ <https://onecompiler.com/>



编程语言众多且流行度在变化

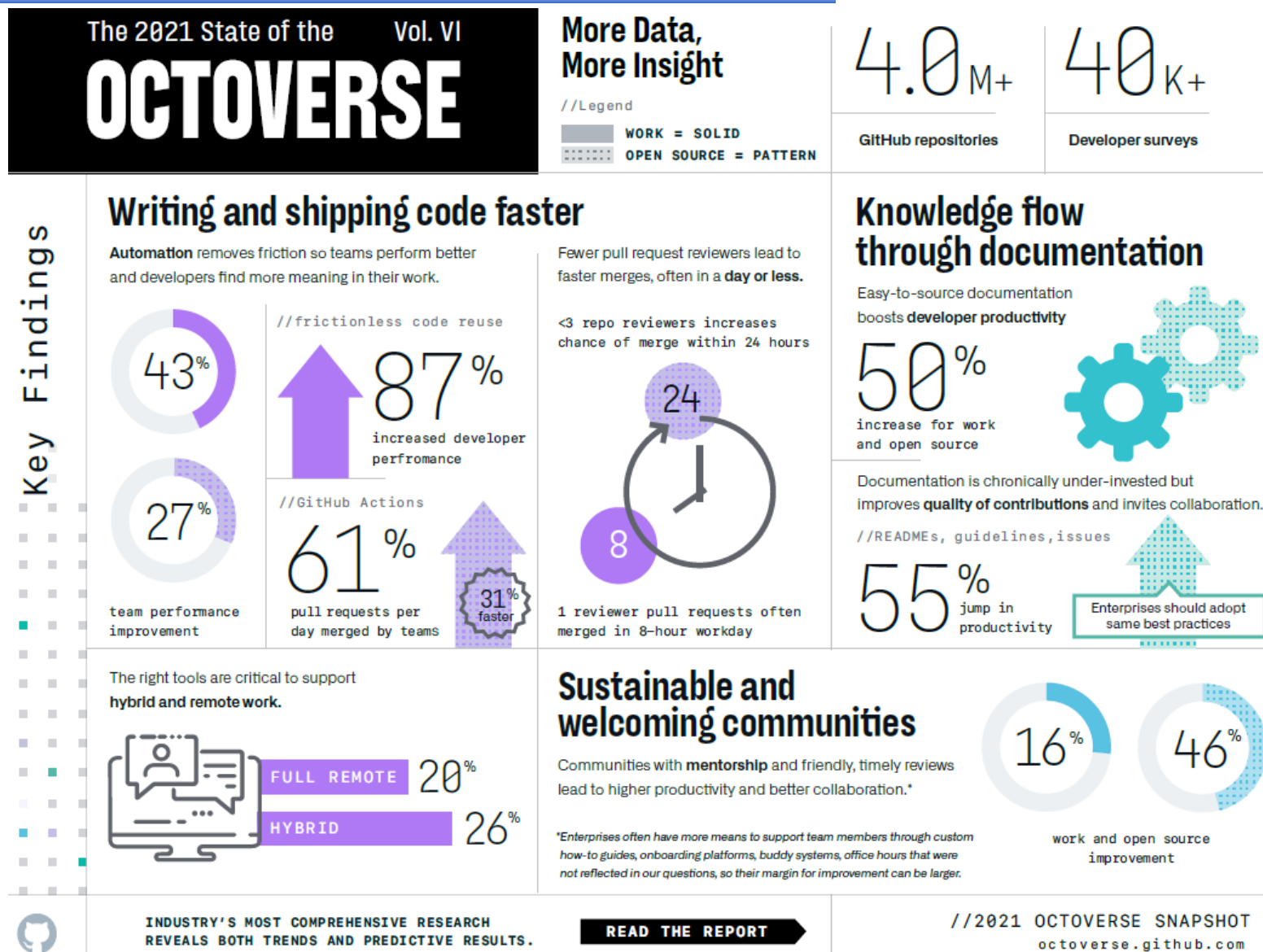
□ GitHub --开源项目涉及370种编程语言 (2019.9)

<https://octoverse.github.com/>





编程语言发展：2021-更快地编码



更快地编写和交付代码

提升开发产能的文档

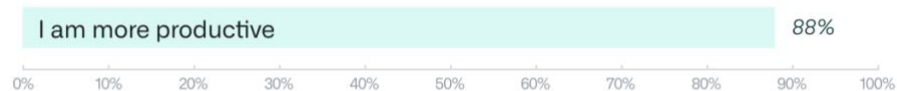


编程语言发展：2022-AI辅助编码

□ Research: quantifying GitHub Copilot's impact on developer productivity and happiness

When using GitHub Copilot...

Perceived Productivity



Satisfaction and Well-being*



Efficiency and Flow*



[ACMQueue202109] [The SPACE of Developer Productivity](#)

We recruited
95
developers, and split them randomly into two groups.

We gave them the task of writing a web server in JavaScript

45 Used
GitHub Copilot

78%
finished

1 hour, 11 minutes
average to complete the task

71 minutes | that's 55% less time!

50 Did not use
GitHub Copilot

70%
finished

2 hours, 41 minutes
average to complete the task

161 minutes

Results are statistically significant ($P=0.017$) and the 95% confidence interval is [21%, 89%]

an evaluation with 24 students,

Google's internal assessment of
ML-enhanced code completion

[ACM Queue2021] The SPACE of Developer Productivity

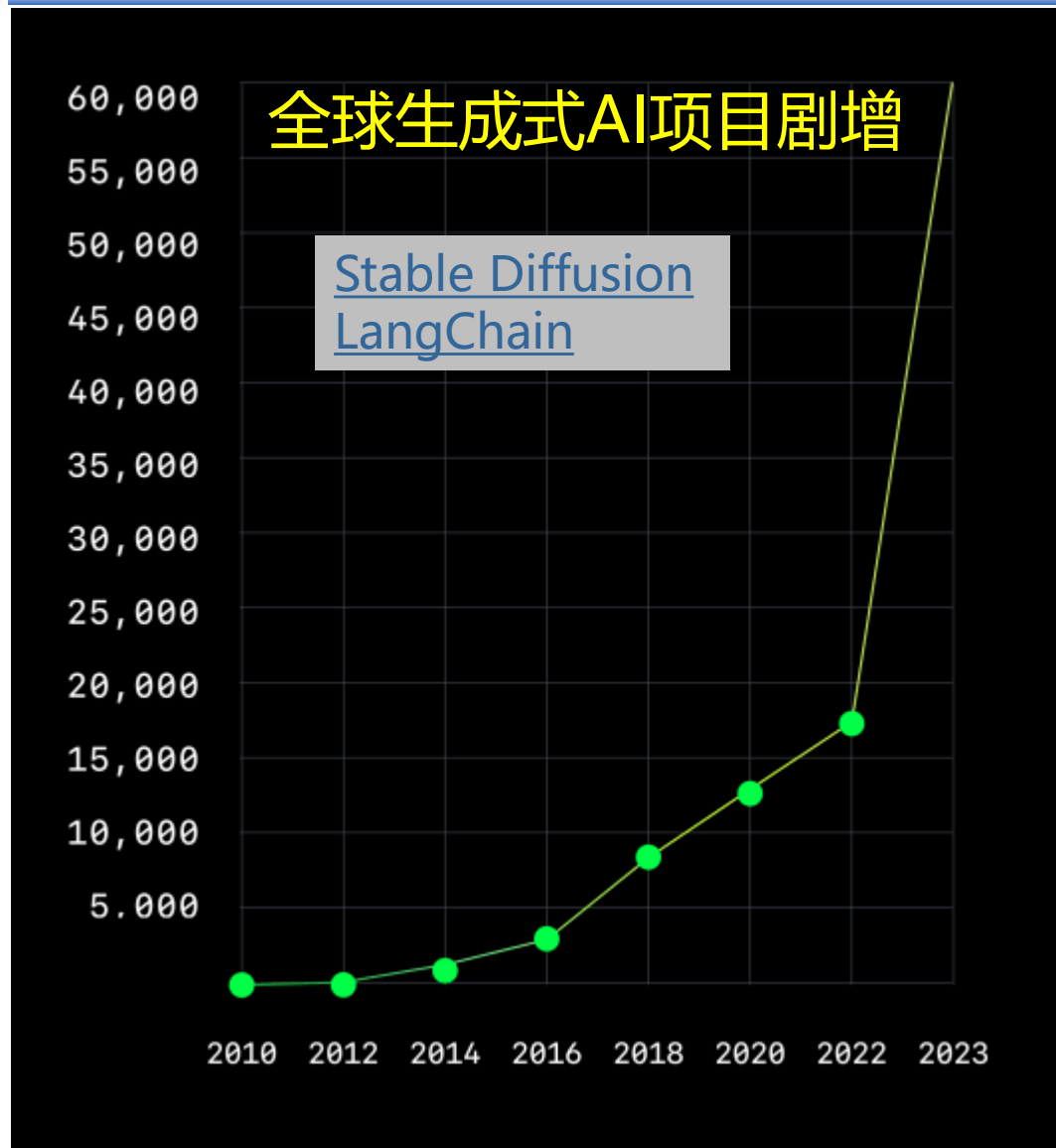
FIGURE 1: EXAMPLE METRICS

LEVEL	SATISFACTION & WELL-BEING How fulfilled, happy, and healthy one is	PERFORMANCE An outcome of a process	ACTIVITY The count of actions or outputs	COMMUNICATION & COLLABORATION How people talk and work together	EFFICIENCY & FLOW Doing work with minimal delays or interruptions
INDIVIDUAL One person	<ul style="list-style-type: none"> * Developer satisfaction * Retention[†] * Satisfaction with code reviews assigned * Perception of code reviews 	<ul style="list-style-type: none"> * Code review velocity 	<ul style="list-style-type: none"> * Number of code reviews completed * Coding time * # Commits * Lines of code[†] 	<ul style="list-style-type: none"> * Code review score (quality or thoughtfulness) * PR merge times * Quality of meetings[†] * Knowledge sharing, discoverability (quality of documentation) 	<ul style="list-style-type: none"> * Code review timing * Productivity perception * Lack of interruptions
TEAM OR GROUP People that work together	<ul style="list-style-type: none"> * Developer satisfaction * Retention[†] 	<ul style="list-style-type: none"> * Code review velocity * Story points shipped[†] 	<ul style="list-style-type: none"> * # Story points completed[†] 	<ul style="list-style-type: none"> * PR merge times * Quality of meetings[†] * Knowledge sharing or discoverability (quality of documentation) 	<ul style="list-style-type: none"> * Code review timing * Handoffs
SYSTEM End-to-end work through a system (like a development pipeline)	<ul style="list-style-type: none"> * Satisfaction with engineering system (e.g., CI/CD pipeline) 	<ul style="list-style-type: none"> * Code review velocity * Code review [acceptance rate] * Customer satisfaction * Reliability (uptime) 	<ul style="list-style-type: none"> * Frequency of deployments 	<ul style="list-style-type: none"> * Knowledge sharing, discoverability (quality of documentation) 	<ul style="list-style-type: none"> * Code review timing * Velocity/flow through the system

[†] Use these metrics with (even more) caution — they can proxy more things.



编程语言发展2023-生成式AI、类型安全↑



□ 类型安全

■ TypeScript (2012~) 渐进类型

首次超越Java

第三受欢迎的语言

其用户群增长了 37%

集语言、类型检查器、编译器和语言服务于一体

■ Rust (2009~) 所有权

[用于系统编程及纳入Linux内核的评论](#)

年使用增长率为 40%

被 2023 年 Stack Overflow 开发者调查评为最受推崇的语言



编程语言众多且流行度在变化











<https://octoverse.github.com/>

□ GitHub --开源项目涉及370种编程语言 (2019.9)

□ TIOBE

<https://www.tiobe.com/tiobe-index/>

编程语言
名人堂
2020, 2018年

Aug 2022	Aug 2021	Change	Programming Language
1	2	▲	 Python
2	1	▼	 C
3	3		 Java
4	4		 C++
5	5		 C#
6	6		 Visual Basic
7	7		 JavaScript
8	9	▲	 Assembly language
9	10	▲	 SQL
10	8	▼	 PHP



□ 编程语言自身在不断发展

- **C** **C90, C99, C11**
- **C++** **1998,..., 2011, 14, 17, 20**

□ 新语言不断产生

- **Go (2009), Rust (2010), Elixir (2011), Swift (2014)**

领域特定语言

- **将高阶函数map、reduce等应用于大数据处理**
大数据处理：MapReduce, Hadoop, ...
- **解耦计算的定义与调度实现**
图像处理：Halide (2012), ... → 深度学习：TVM...
- **深度学习编程框架**：TensorFlow → JAX, PyTorch, MindSpore, ...
- **图查询语言**：GQL, Cypher, PGQL, ...



HOPL:

History of Programming Languages

<https://dl.acm.org/conference/hopl>



高阶函数→闭包

高阶函数在现代语言中被越来越多地支持

```
def outer(x):  
    def inner(y):  
        return x + y  
    return inner
```

`outer`是返回函数inner的高阶函数

```
a = outer(2)  
print('function:',a)  
print('result:',a(3))
```

`a`得到函数inner

`a(3)` 调用时要计算 `x+3`

其中`x`是不在inner中定义的非局部变量



引入**闭包closure**:

将 `x=2`作为inner返回值的环境，形成闭包来返回

=> `a(3)` 调用时要计算 `x+3`，可从闭包中获取`x`的值

DSL领域特定语言

□ Halide: 面向图像处理的DSL

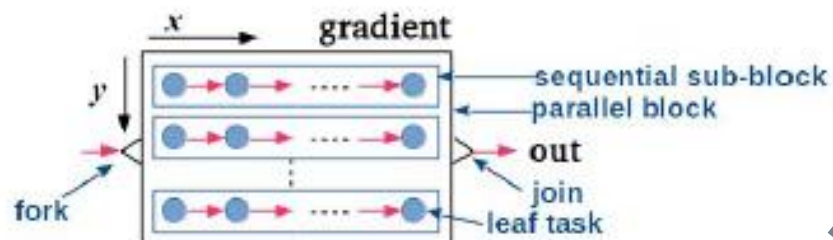
(a) Halide program example

```
Var x, y;  
Func gradient;  
gradient(x, y) = x + y;  
gradient.parallel(y);  
out = gradient.realize(1024, 1024);
```

计算的定义

计算的调度

(b) Scheduled task graph



(c) Intermediate representation

```
alloc gradient[1024][1024]  
parallel for y in 0...1023:  
  for x in 0...1023:  
    gradient[y][x] = x + y
```

(d) Result after lowering parallel loop

```
define task_function(task_num, closure):  
  gradient = unpacking(closure)  
  for x in 0...1023:  
    gradient[task_num][x] = x + task_num  
  
alloc gradient[1024][1024]  
closure = packing(gradient)  
halide_do_par_for(task_function, 0, 1024, closure)
```

计算的定义与调度分离



编程语言的设计

□ 为什么那么多语言？

- 单个语言不能适用所有应用
- 程序员对语言的好坏、如何编程有自己的观点和看法
- 没有评价语言好坏的普遍接受的标准

□ 语言进化之驱动力

- 应用的多样性
- 提高软件开发生产力(productivity)
- 改善软件的安全性、可靠性和可维护性
- 支持并行(parallelism)与并发(concurrency)
- 移动和分发、模块化、多范型

程序语言设计的计算思维

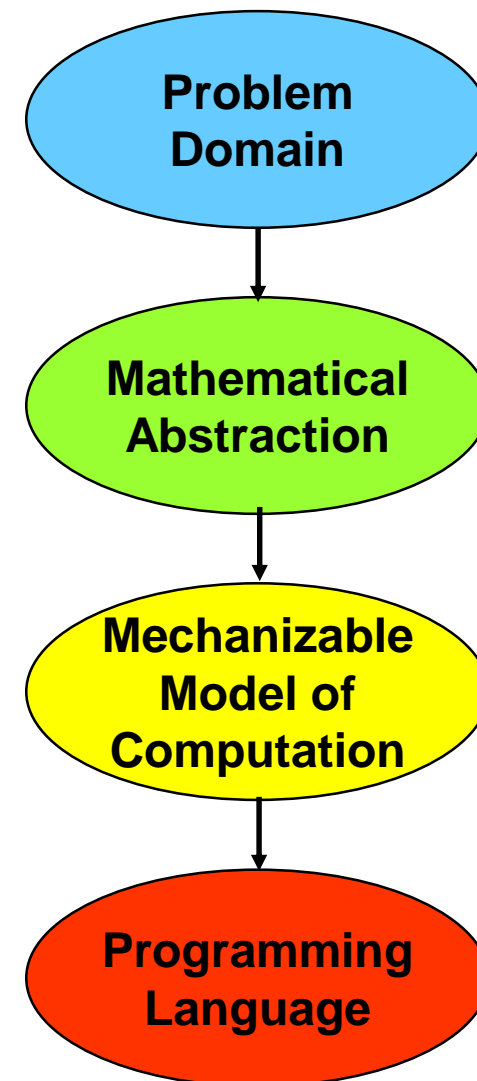
□ 计算思维 (Computational Thinking)



Jeannette M. Wing
Computational Thinking
CACM, vol. 49, no. 3, pp. 33-35, 2006

Computational thinking is a **fundamental skill for everyone**, not just for computer scientists. To reading, writing, and arithmetic, **we should add computational thinking to every child's analytical ability**. Just as the printing press facilitated the spread of the three Rs, what is appropriately incestuous about this vision is that computing and computers facilitate the spread of computational thinking.

□ 语言设计中的计算思维





主要内容

1

编程语言及设计

2

编译器的阶段

3

编译器的作用和形式

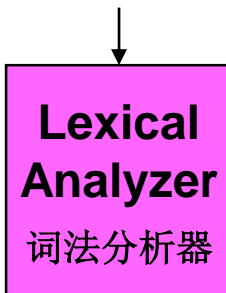
4

编译技术的应用与挑战



编译器的阶段

source
program



Token
Stream
记号流

□ 词法分析：将程序字符流分解为记号
(Token) 序列

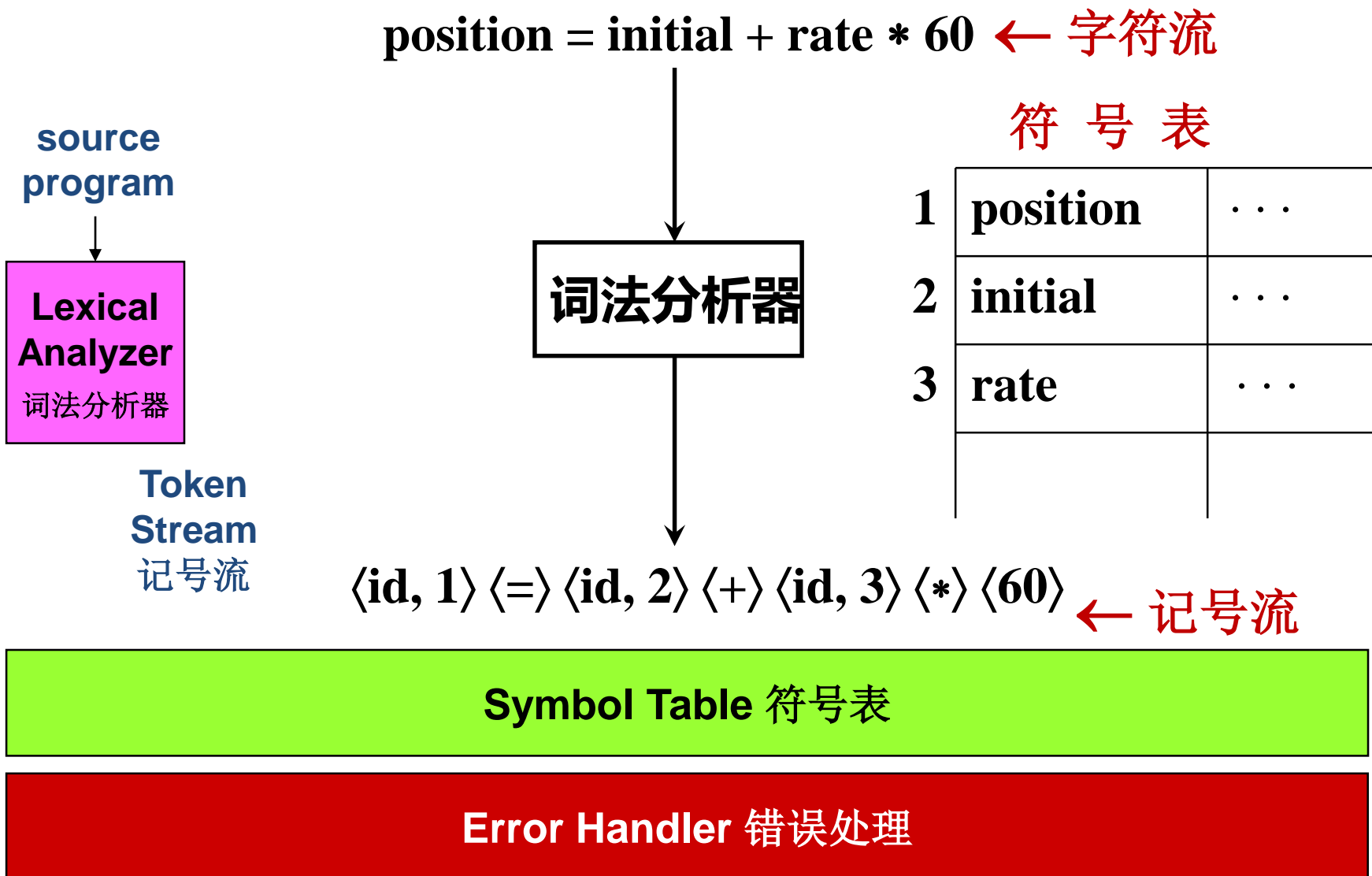
◆ 形式：<token_name, attribute_value>

Symbol Table 符号表

Error Handler 错误处理

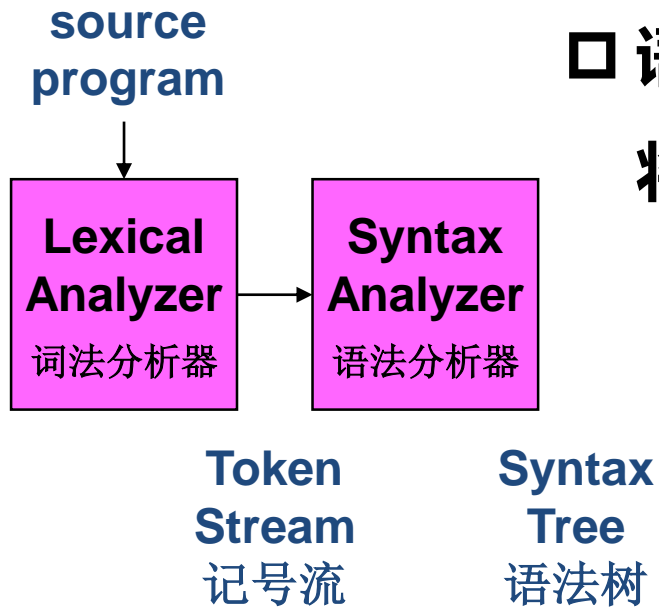


编译器的阶段





编译器的阶段



□ 语法分析：也称解析(Parsing)

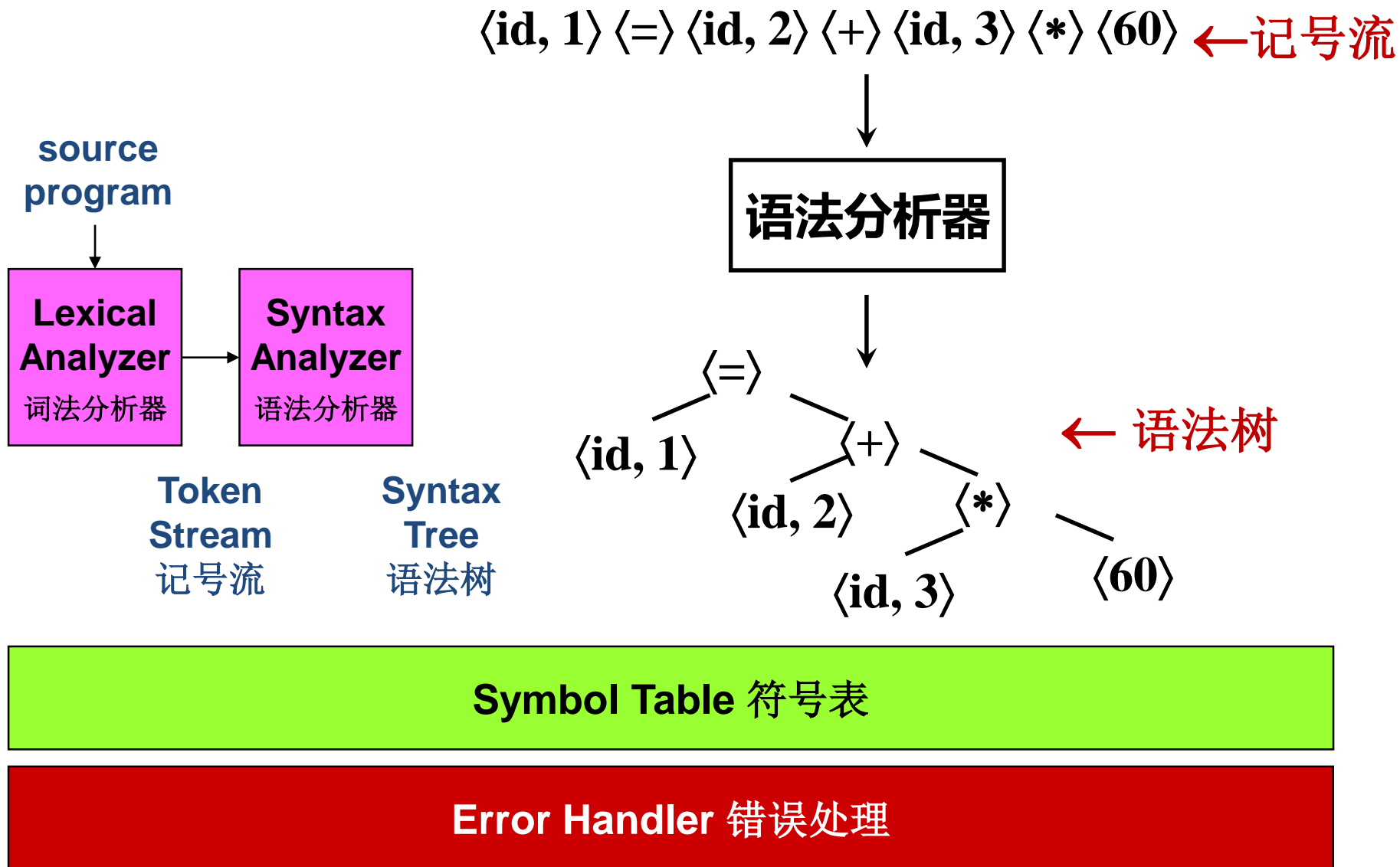
将记号序列解析为语法结构

Symbol Table 符号表

Error Handler 错误处理

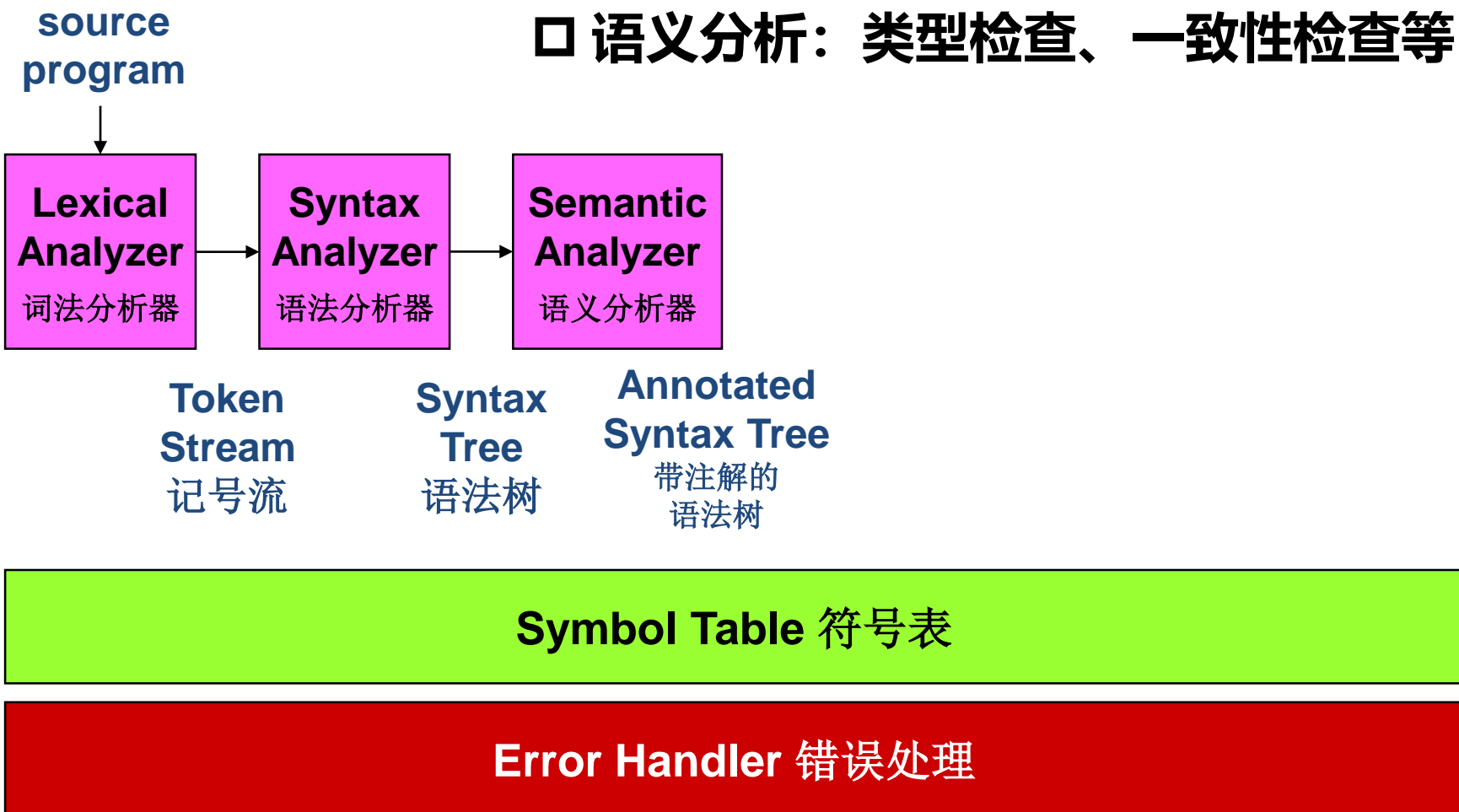


编译器的阶段



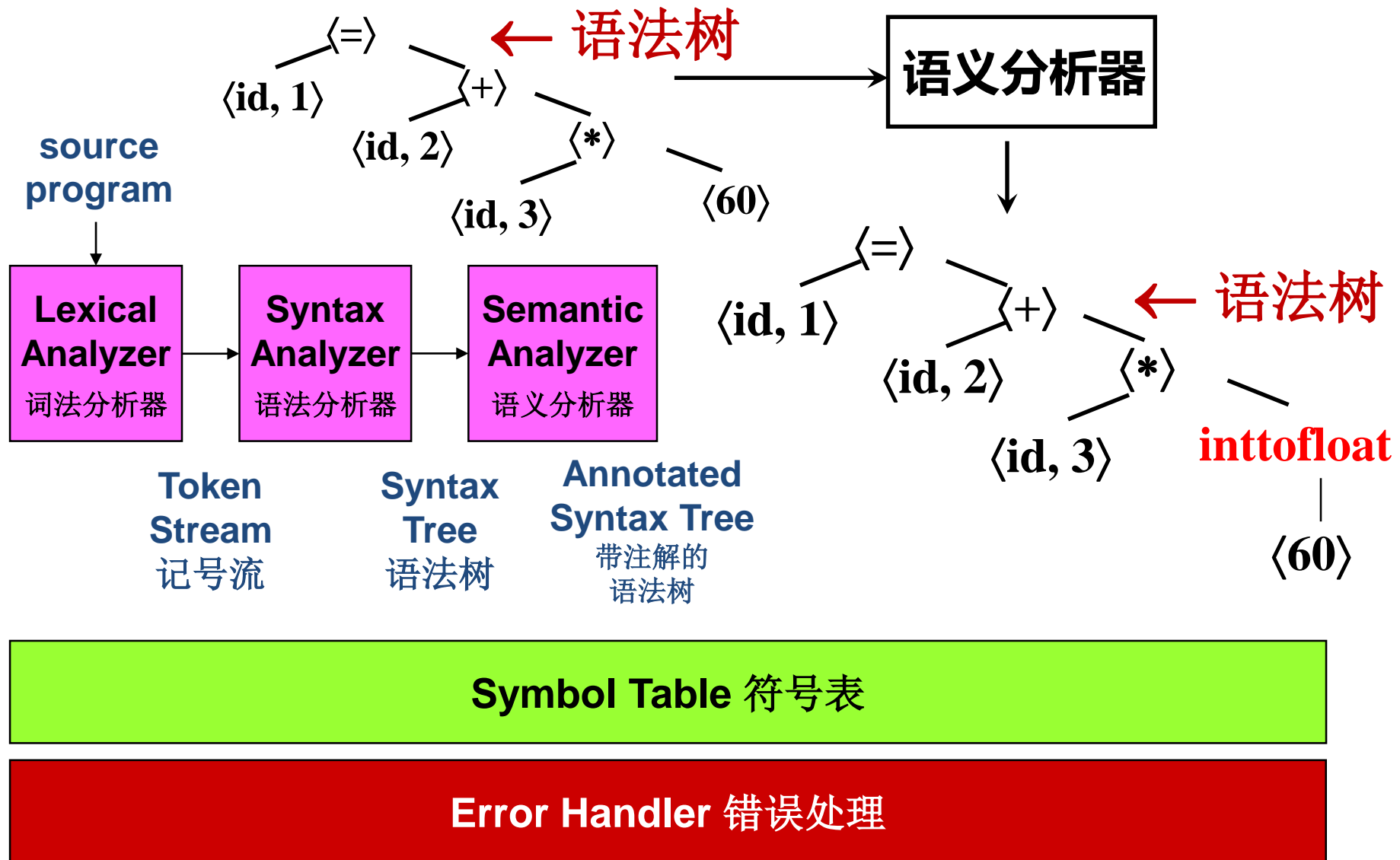


编译器的阶段





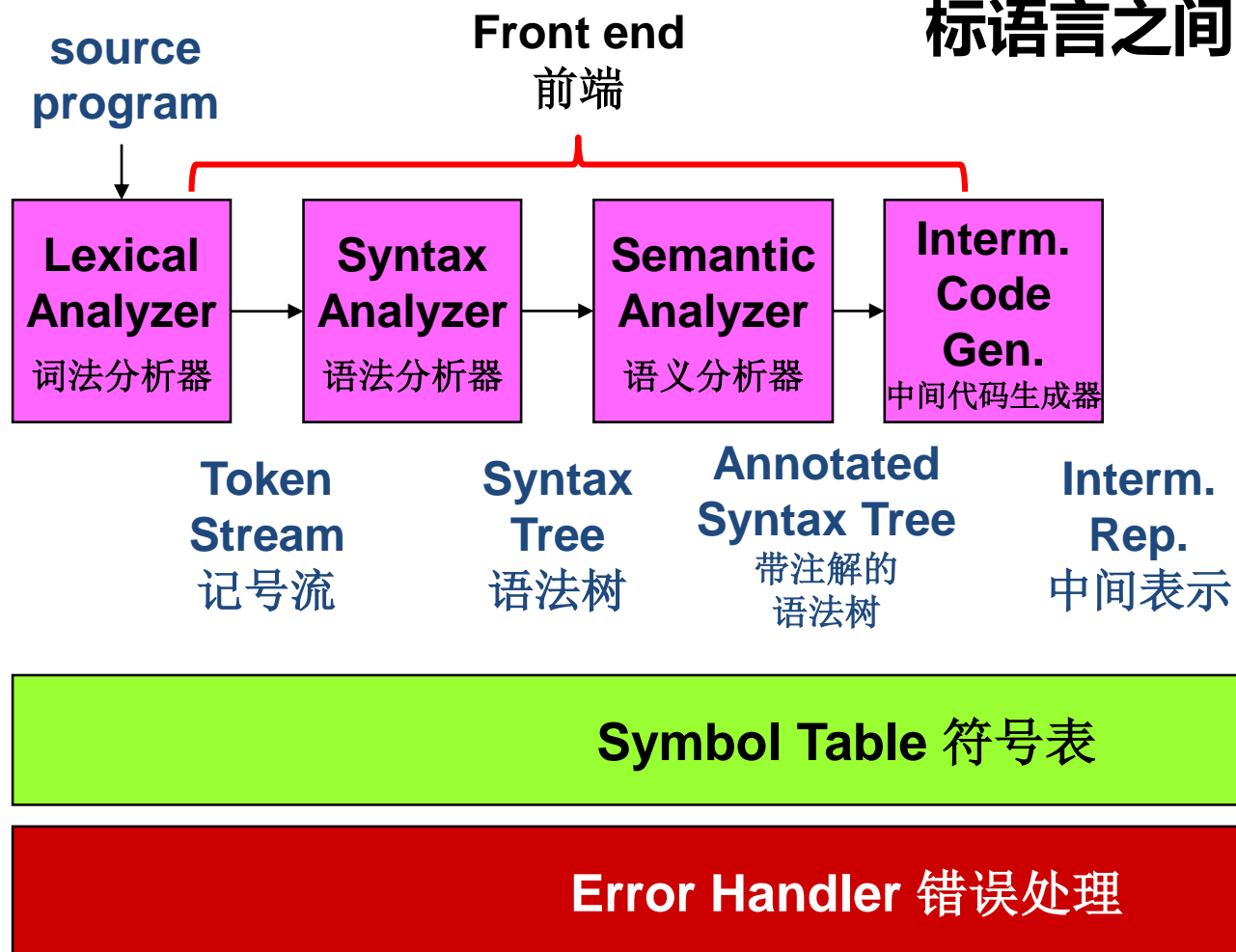
编译器的阶段





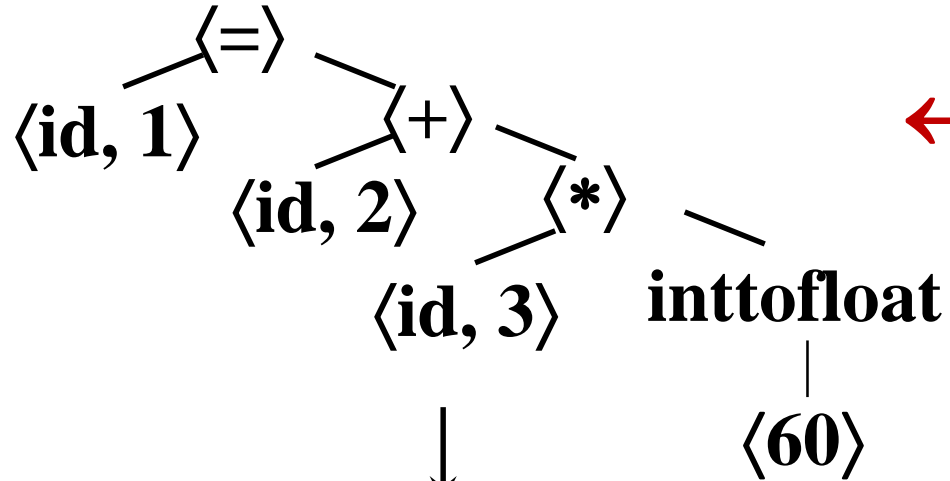
编译器的阶段

□ 中间代码生成：源语言与目标语言之间的桥梁





编译器的阶段



← 语法树

中间代码生成器

t1 = inttofloat(60)
t2 = id3 * t1
t3 = id2 + t2
id1 = t3

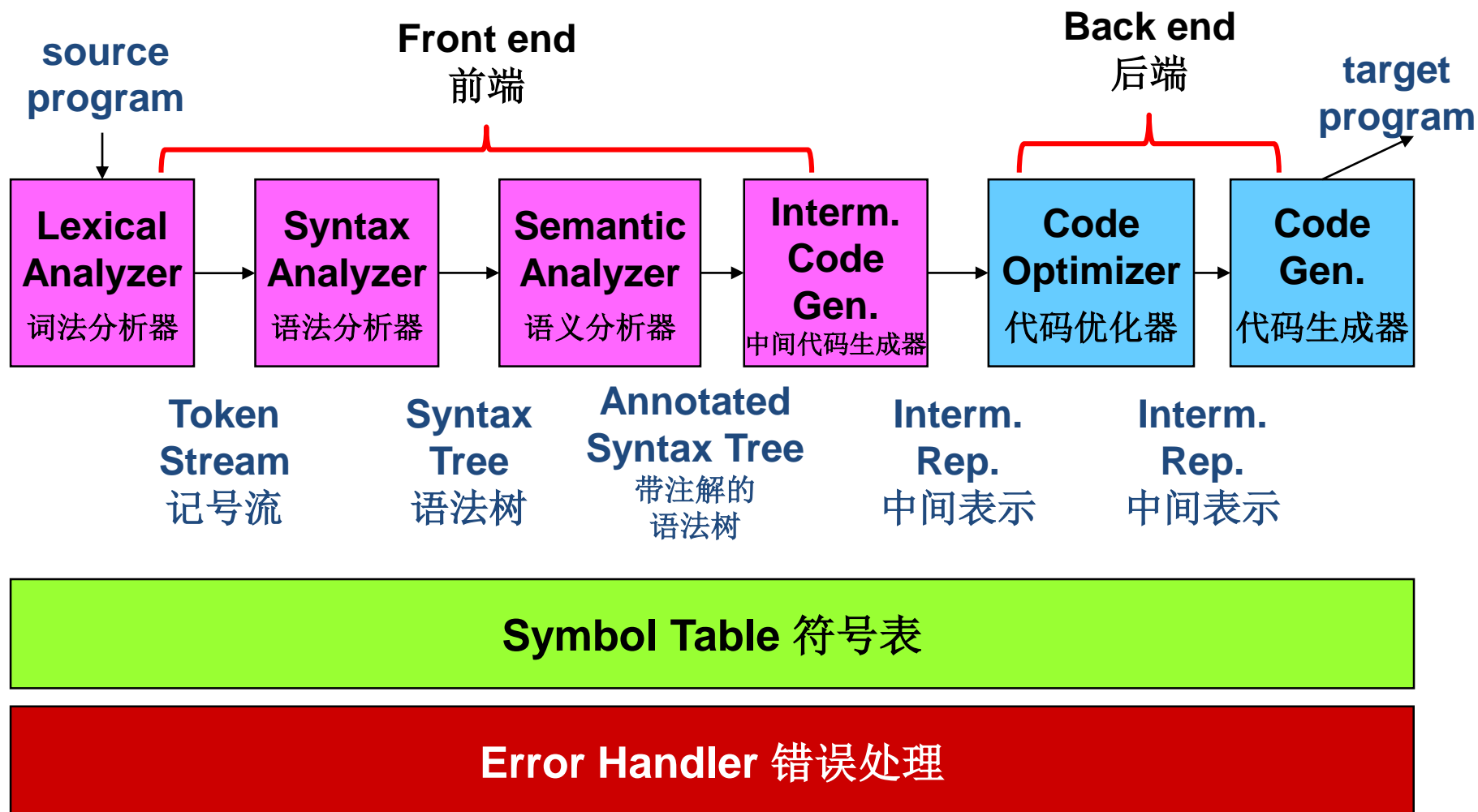
← 三地址中间代码

符号表

1	position	...
2	initial	...
3	rate	...



编译器的阶段





代码优化

- 机器无关的优化、机器相关的优化
- 降低执行时间，减少能耗、资源消耗等

t1 = inttofloat(60) ← 三地址中间代码

t2 = id3 * t1

t3 = id2 + t2

id1 = t3



代码优化器



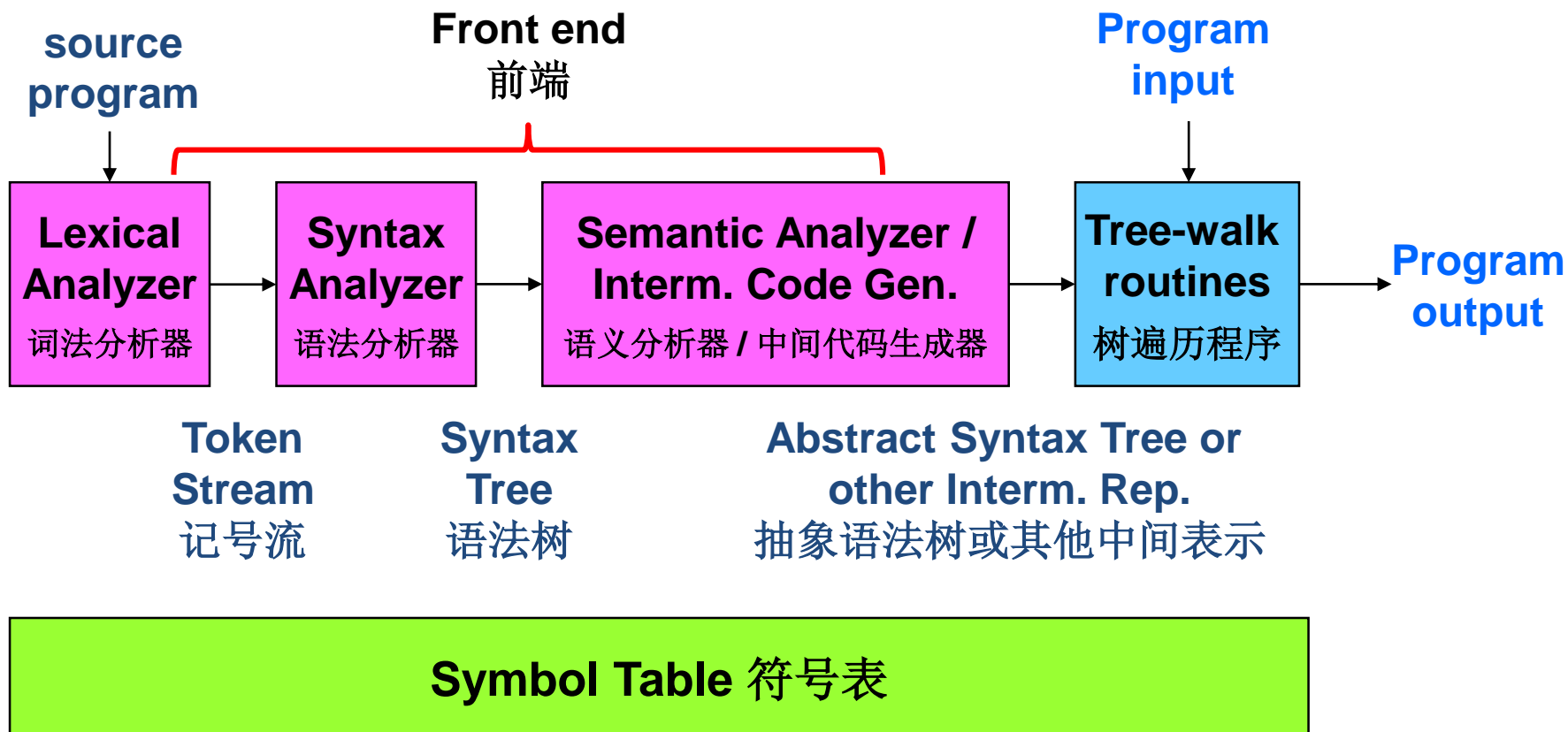
t1 = id3 * 60.0

id1 = id2 + t1

← 三地址中间代码

符号表

1	position	...
2	initial	...
3	rate	...





主要内容

1

编程语言及设计

2

编译器的阶段

3

编译器的作用和形式

4

编译技术的应用与挑战



编译器的作用

□ 编程语言

□ 目标机器语言

□ 编译器（编译系统）

■ 主流的开源编译基础设施：[GCC](#)、[Clang/LLVM](#)

```
[root@host ~]# gcc helloworld.c -o helloworld
```

```
[root@host ~]# ./helloworld
```

```
hello, world!
```

```
#include <stdio.h>
int main()
{
    printf("hello, world!\n");
}

/* helloworld.c */
```

注意：gcc是驱动程序
(根据命令行参数调用相应的处理程序)



编译器的作用

□ 翻译

- 支持高层的编程抽象
- 支持底层的硬件体系结构

□ 优化

- 更快的执行速度
- 更少的空间

□ 分析

- 程序理解
- Safety: 自身的稳定状态, 功能正确
- Security: 免受外部伤害



举例：性能与安全

```
for (i=0; i<n; i++) a[i] = 1;
```

```
pend = a+n;
```

```
for (p=a; p<pend; p++) *p = 1;
```

哪个更快，Why?

```
foo (char * s)
```

```
{
```

```
    char buf[32];
```

```
    strcpy (buf, s);
```

```
}
```

调用foo()会如何?



举例：性能与安全

```
for (i=0; i<n; i++) a[i] = 1;
```



```
pend = a+n;  
for (p=a; p<pend; p++) *p = 1;
```

哪个更快，Why?

```
foo (char * s)  
{  
    char buf[32];  
    strcpy (buf, s);  
}
```

调用foo()会如何?

若s指向的串的长度超出31，则复制时会超出buf数组的有效区域

目标语言

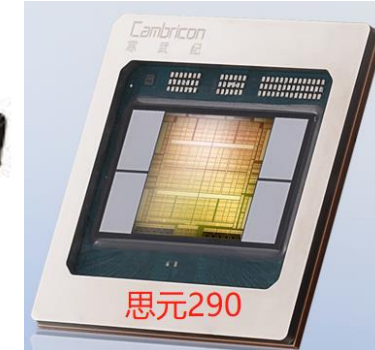
- 另一种编程语言
- CISCs（复杂指令集）：[x86](#)、[IA64](#)、...
- RISCs（精简指令集）：[MIPS](#)、[ARM](#)、[LoongArch指令集](#)、...
- 多核/众核
- GPUs：[CUDA](#)、[OpenCL](#)
- FPGAs
- 异构编程 [SYCL](#)
- 量子计算机
- TPU, NPU
- ...



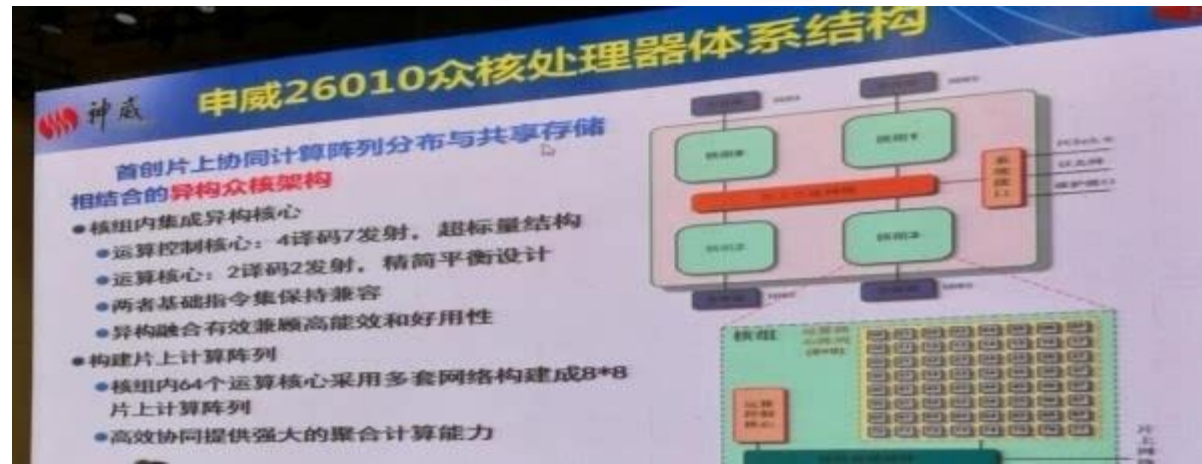
龙芯3A5000



龙芯3A5000计算机

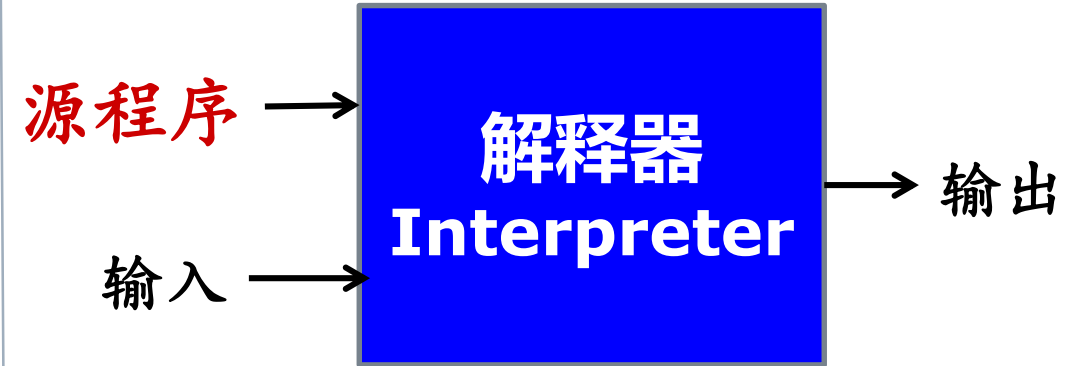
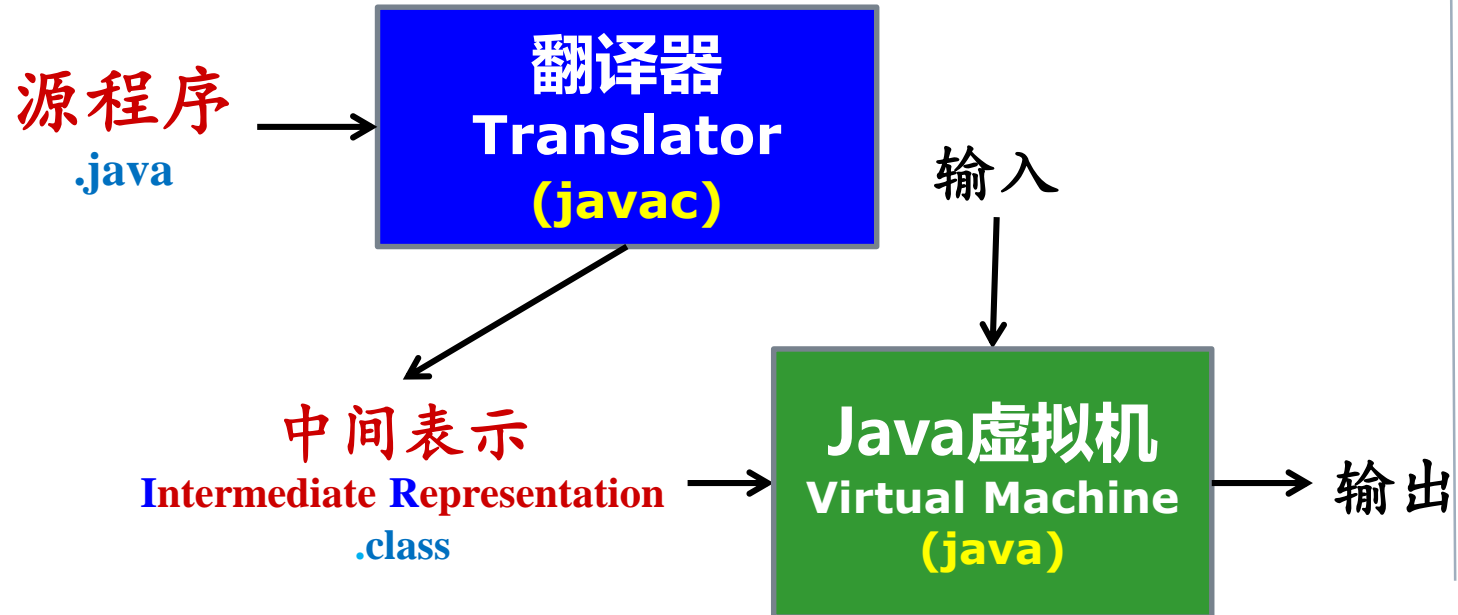
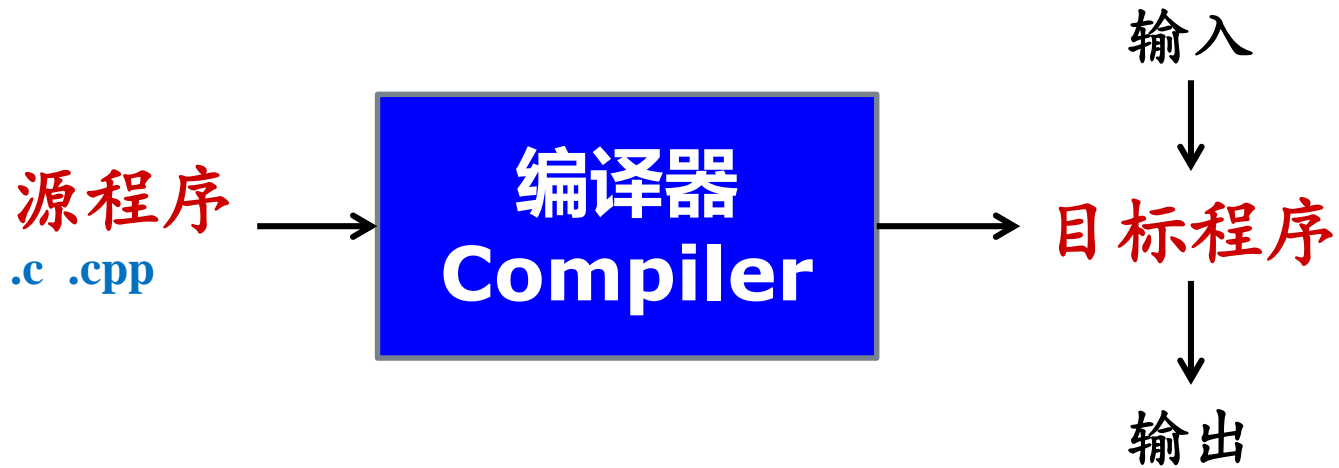


思元290





编译器的形式



直接在输入上执行源程序
如Python等脚本语言

执行效率低，但容易编写



编译器的其他形式

□ 交叉编译器 (Cross compiler)

- 在一个平台上生成另一个平台上的代码

PC → **arm-linux-gcc** → ARM

□ 增量编译器 (Incremental compiler)

- 以增量地编译源程序,只编译修改的部分,如 [Freeline](#)

□ 即时编译器 (Just-in-time compiler)

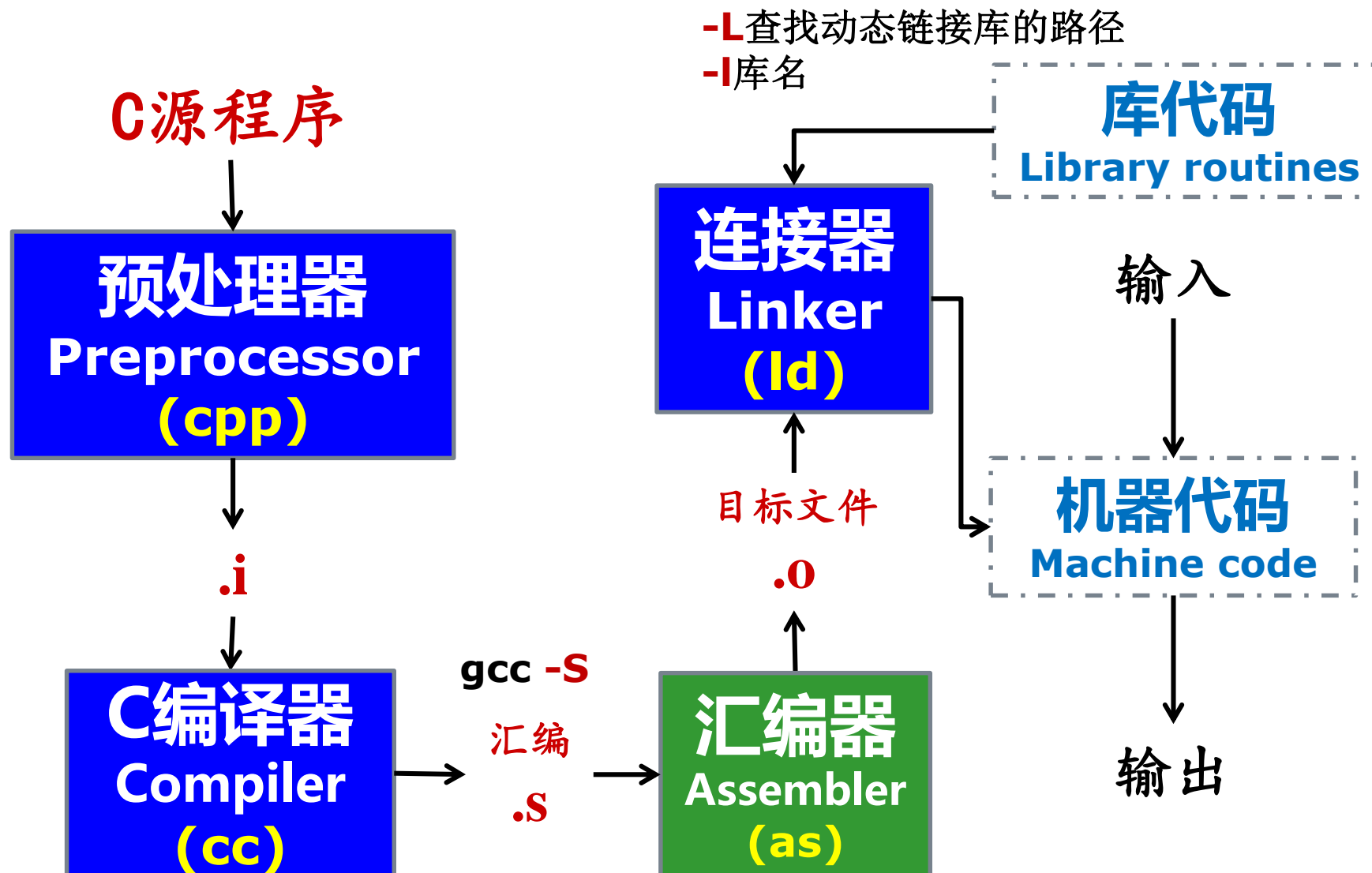
- 在运行时对IR中每个被调用的方法进行编译,得到目标机器的本地代码,如 Java VM 中的即时编译器

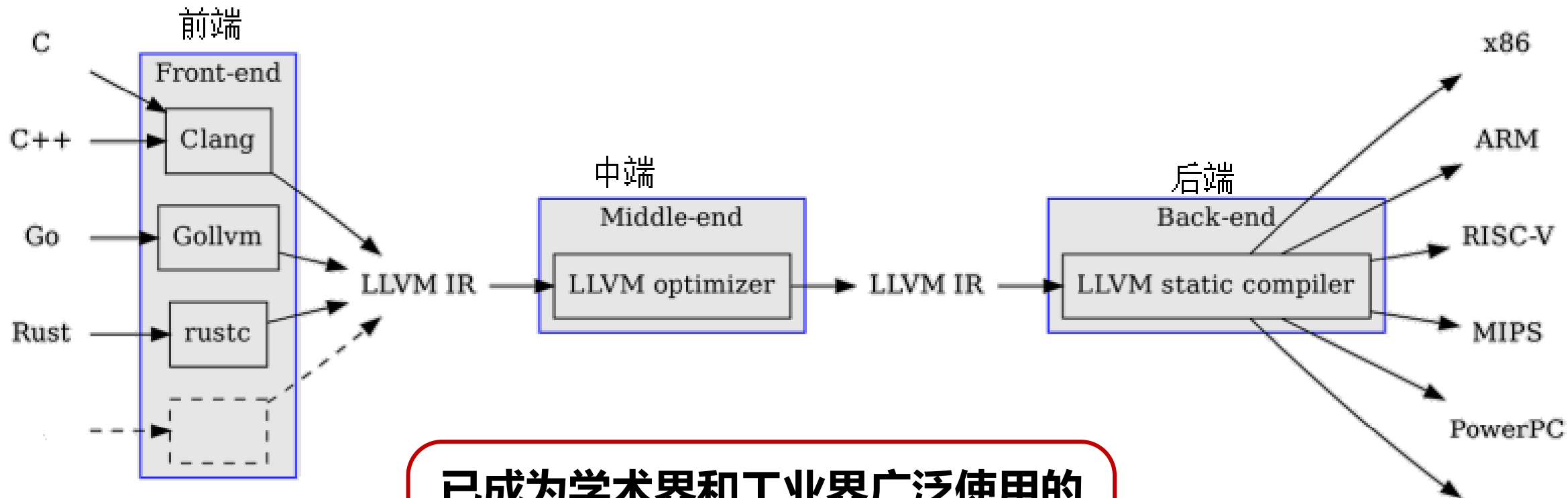
□ 提前编译器 (Ahead-of-time compiler)

- 在程序执行之前将IR翻成本地码,如 ART中的AOT



C 编译器





已成为学术界和工业界广泛使用的
编译器基础设施



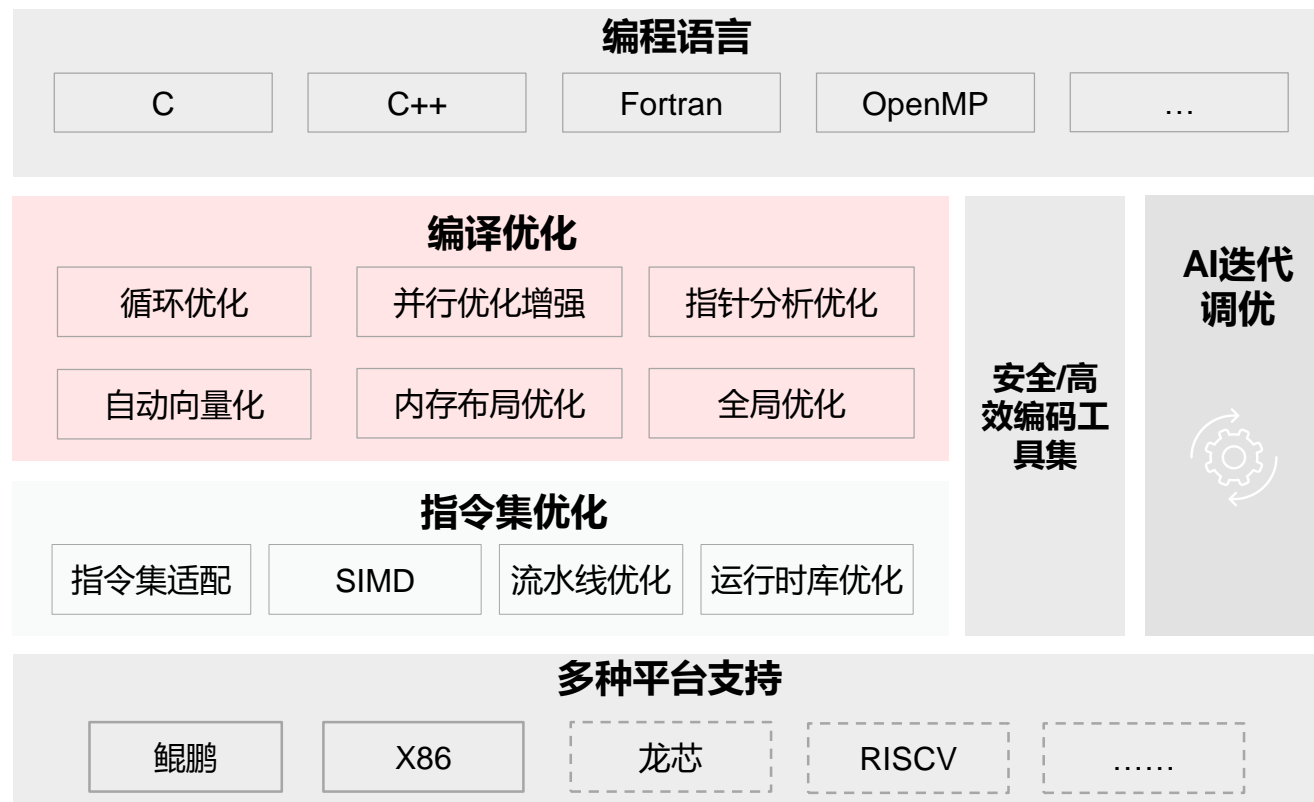
国产化：毕昇编译器

□ 基于LLVM扩展

- 支持C/C++/Fortran编程语言
- 增强中端编译优化技术
- 支持鲲鹏920、X86-64等硬件

□ 打造高性能、高可信、易扩展的编译工具链

- 性能/代码体积优化选项
- 多样化浮点精度选项/模式调优
- 静态检查工具，重构工具
- 支持AI迭代调优，自动优化编译配置
- 针对通用处理器架构的各类高性能编译优化技术





毕昇编译器历史沿革

编译器团队成立

- 第一批海内外研究人员加入

2009, 从零构建

无线DSP编译器业界领先

- DSP编译器性能超越标杆XCC, 基站竞争力超越E
- CPU/DSP编译器规模商用, 在网规模超xxw

2013, 规模商用

ARM64编译器竞争力领先

- 自研ARM C/C++编译器实现ARM64 领域竞争力领先, 并在华为公司多个场景商用;
- DSP编译器支撑基站、控制器等海量发货xx万套, 服务全球xx亿用户

2016, 业界领先

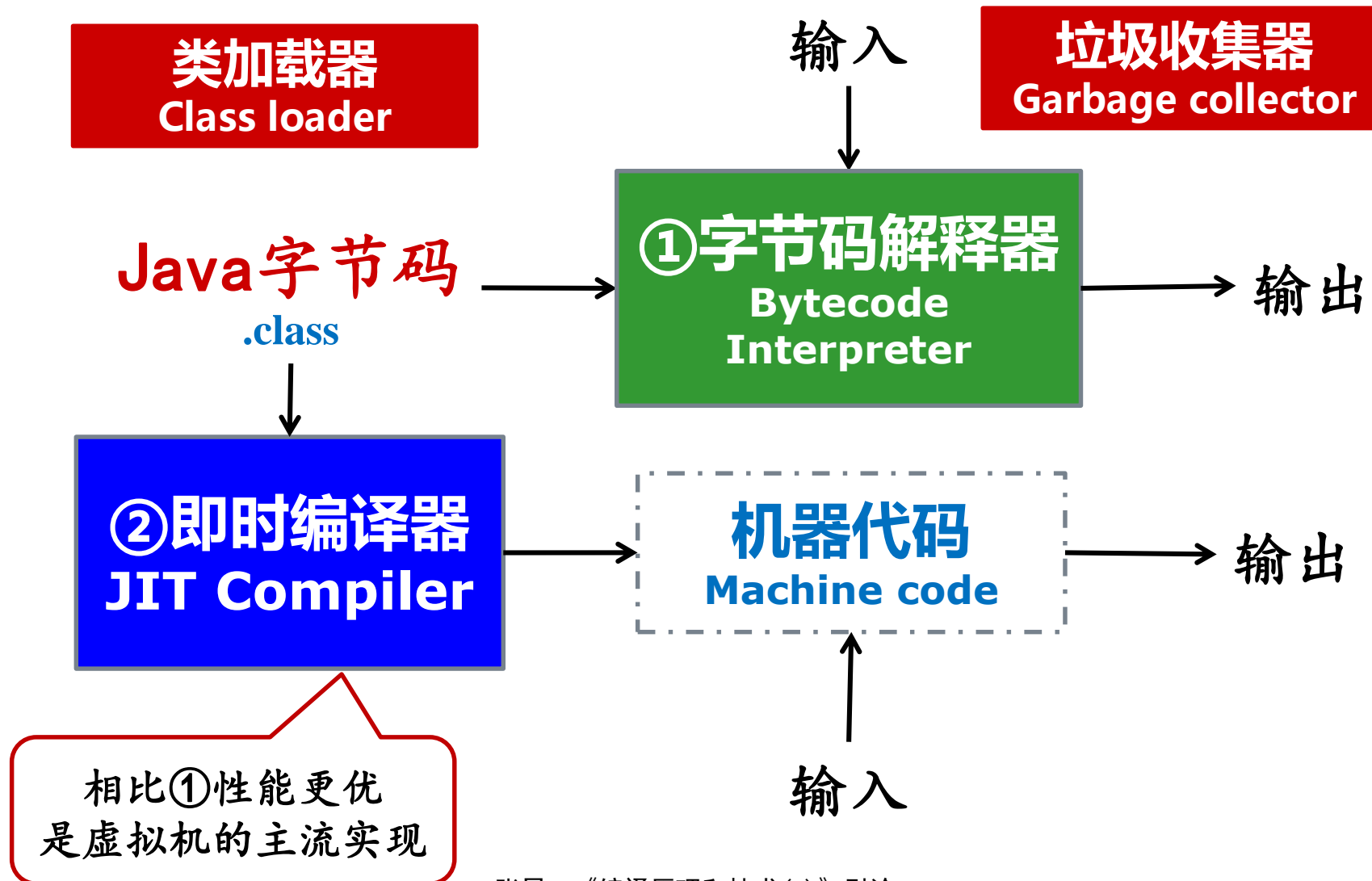
计算/AI等新领域竞争力突破

- 发布毕昇编译器, 实现鲲鹏原生性能提升和多样算力融合优化
- 发布二进制翻译工具ExaGear, 实现x86到ARM生态平滑迁移
- 昇腾编译器完整支持ISA6.3与V100全系列芯片, 助力Atlas集群挺进秒级训练

2019, 持续领航



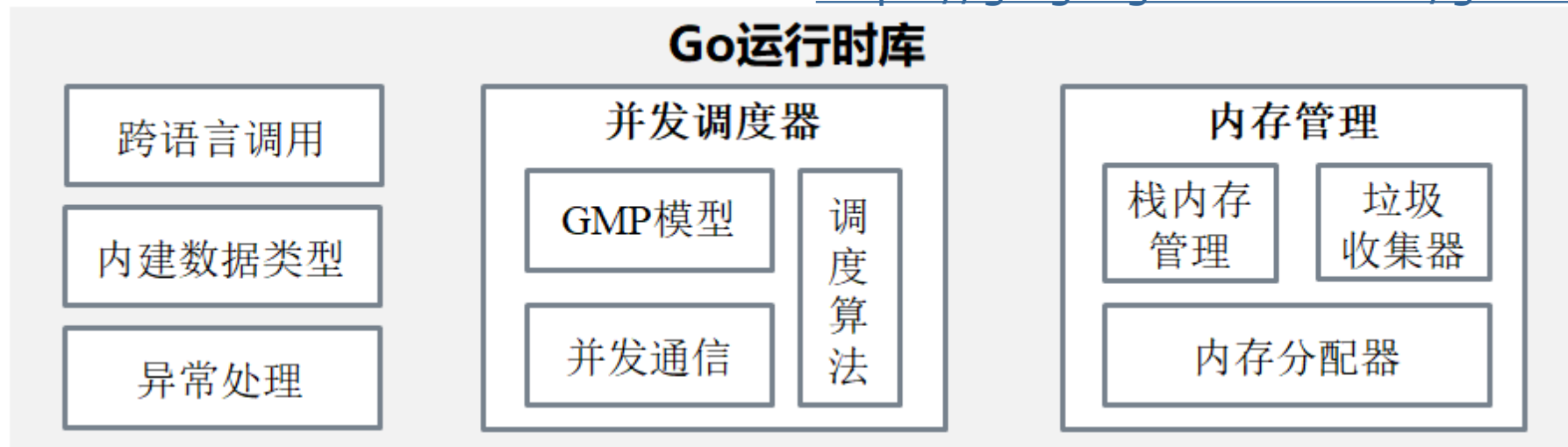
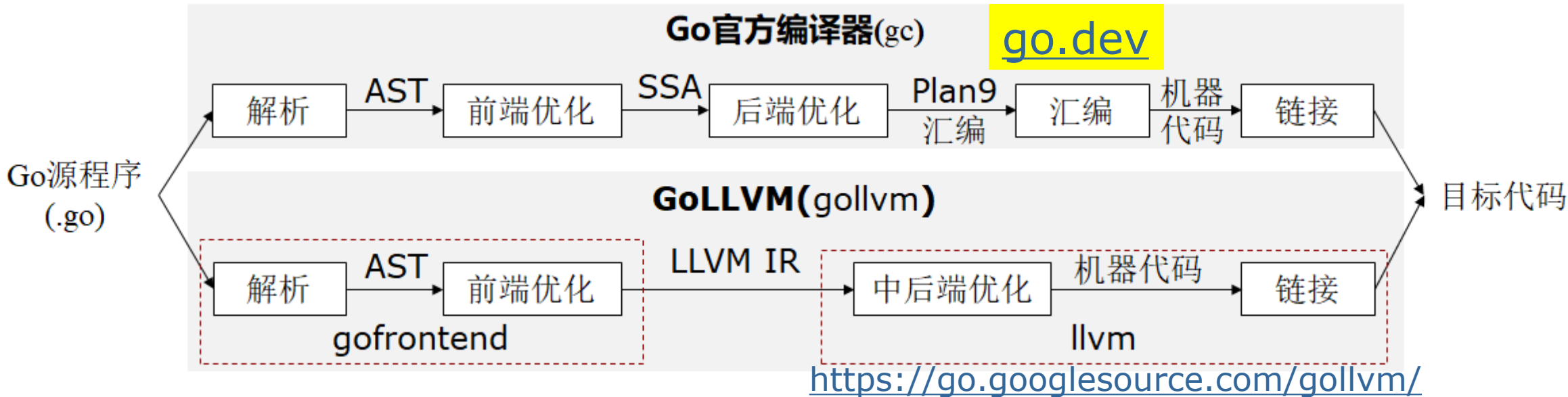
Java虚拟机





Go语言编译器

带GC的编译型语言



易于编程
快速编译
高效运行

<https://pkg.go.dev/runtime>



主要内容

- 1 编程语言及设计
- 2 编译器及形式
- 3 编译器的阶段
- 4 **编译技术的应用与挑战**



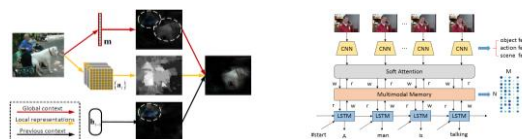
人工智能应用及深度学习框架

无人驾驶系统的软件栈

应用层



模型层



框架层



高层
中间表示

NNVM
Relay



编译

算子
实现层



优化

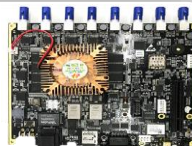
硬件层



Nvidia Drive
PX2



Nvidia Jetson
Nano



地平线“征途2.0”

国产系统软件与硬件



CANN

Compute
Architecture
for Neural
Networks

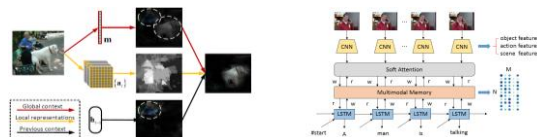


无人驾驶系统的软件栈

应用层



模型层



框架层



高层
中间表示

NNVM
Relay



编译

算子
实现层



优化

硬件层



Nvidia Drive PX2



Nvidia Jetson Nano



地平线 "征程2.0"

国产系统软件与硬件

科学计算应用/人工智能应用

编程框架/库/中间件

SWTVM SWTensorFlow SWPyTorch SWMind 应用平台基础框架

人工智能算子库 基础数学库 SWBlas SWSparse Shentu

编程语言

C C++ Fortran Python MPI OpenACC 并行C

基础组件

众核基础编译器 Athread运行时 动态运行支撑

支持SACA的神威平台

"神威"系列超级计算机

"神威"众核服务器



多语言软件及优化

- Python + C / C++
- Java + C / C++
- JavaScript + C / C++
- Go + C / C++
- Rust + C/C++

跨语言程序分析

类型推断、跨语言程序调用图、
资源分析与管理、信息流分析等

安全可靠、性能极致



关键科学问题

□ 语言定义

- 如何抽象和形式化
- 如何推陈出新



正规式
上下文无关文法
类型系统

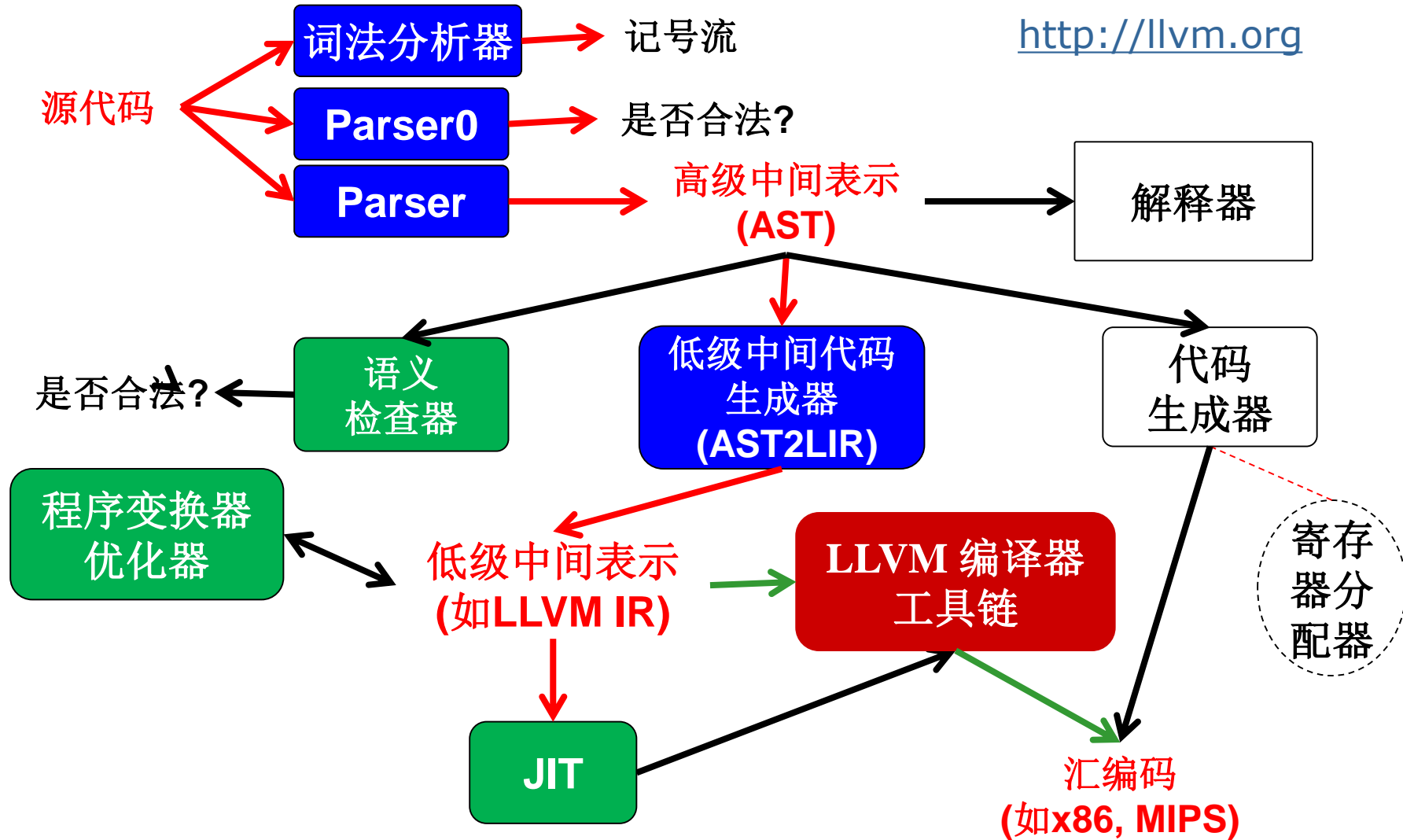
□ 应对不断发展的应用和硬件

- 发挥硬件及指令集优势的代码生成
- 软硬件协同设计
- 增强软硬件系统的健壮性

中间表示设计与生成
数据流/控制流分析
代码生成与优化



基础实验的考虑





我听到的会忘掉，
我看到的能记住，
我做过的才真正明白。