

计算图与 优化与执行



ZOMI



关于本内容

1. 内容背景

- 计算图基础介绍

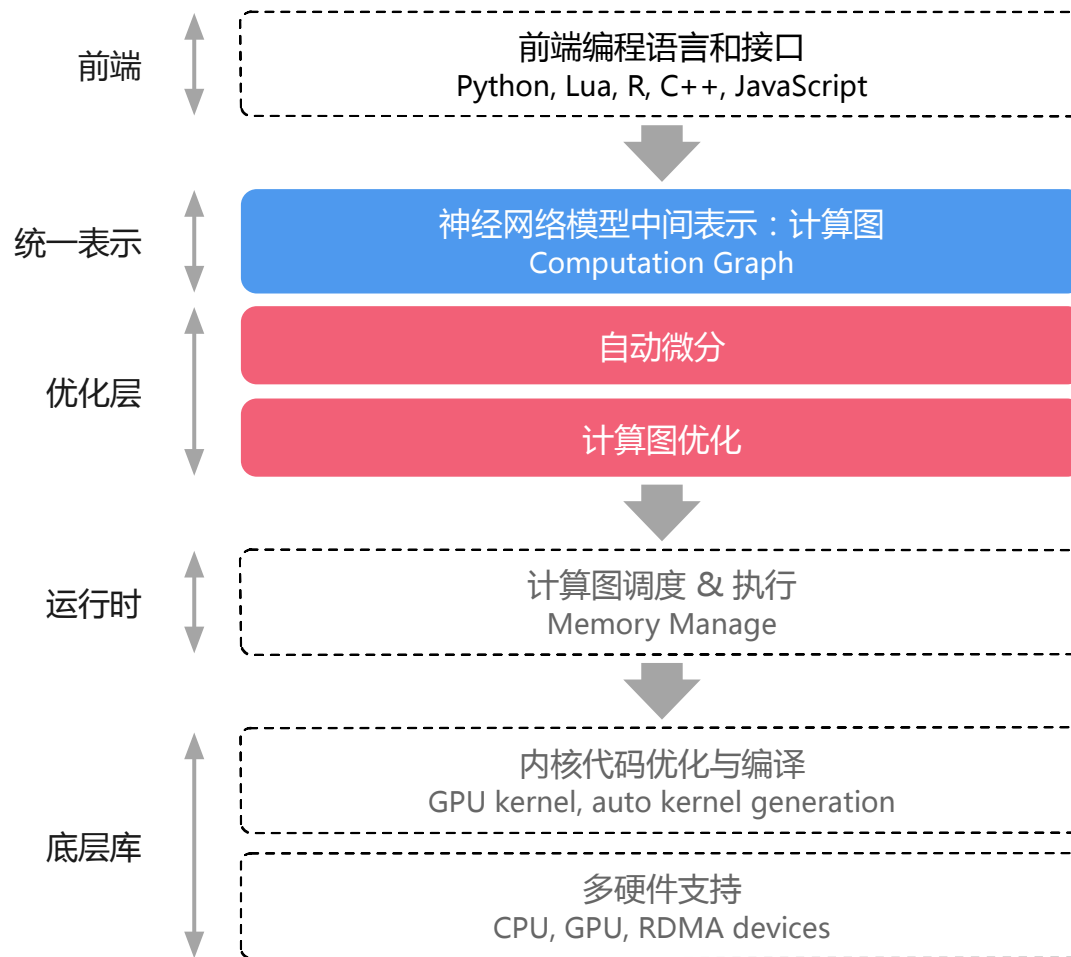
2. 具体内容

- 计算图（数据流图）：AI系统化问题 – 计算图的提出
- 计算图和自动微分：深度学习与微分 - 回顾自动微分 – 计算图表达自动微分
- 图的调度和执行：图优化 – 图调度与执行
- 图与控制流：控制流表达 – 动态图控制方式 – 静态图控制方式 – 动静统一
- 计算图的挑战与未来

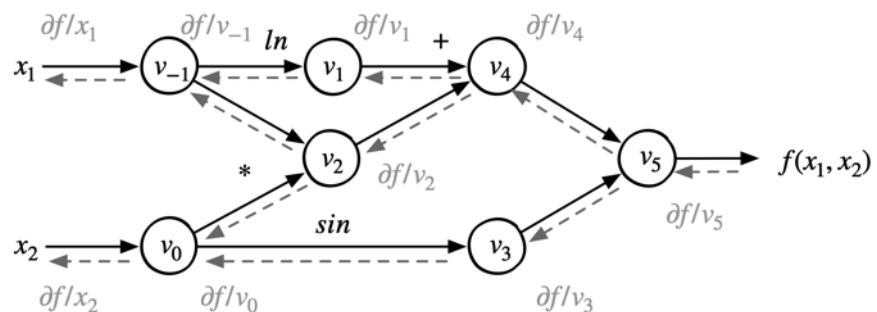
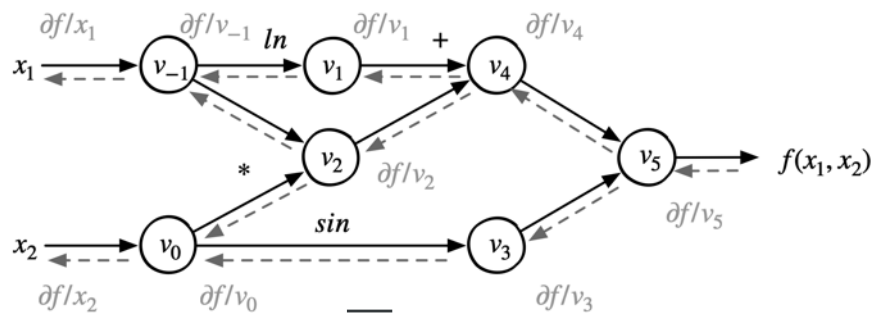
图优化

- 计算图的出现允许 AI 框架在执行前看到深度学习模型定义全局信息
- 计算图作为 AI 框架中的高层中间表示，可以通过图优化 Pass 去化简计算图或提高执行效率

利用反向微分计算梯度通常实现为数据流图上的一个优化 Pass



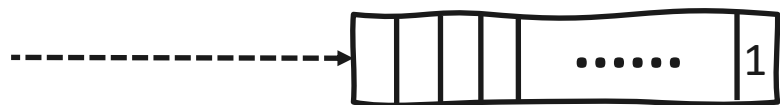
图优化



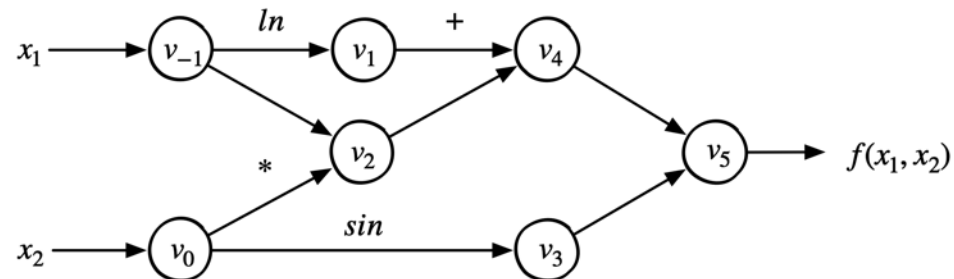
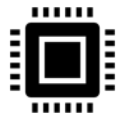
计算图的调度与执行

根据依赖关系，依次调度运行代码

1. $\text{Log}(v_{-1}, 2) \rightarrow v_1$
2. $\text{Mul}(v_{-1}, v_0) \rightarrow v_2$
3. $\text{Sin}(v_0) \rightarrow v_3$
4. $\text{Add}(v_1, v_2) \rightarrow v_4$
5. $\text{Sub}(v_4, v_3) \rightarrow v_5$
6. $\text{Sub_grad}(v_5, v_5_delta) \rightarrow v_4_delta$
7.



Operators



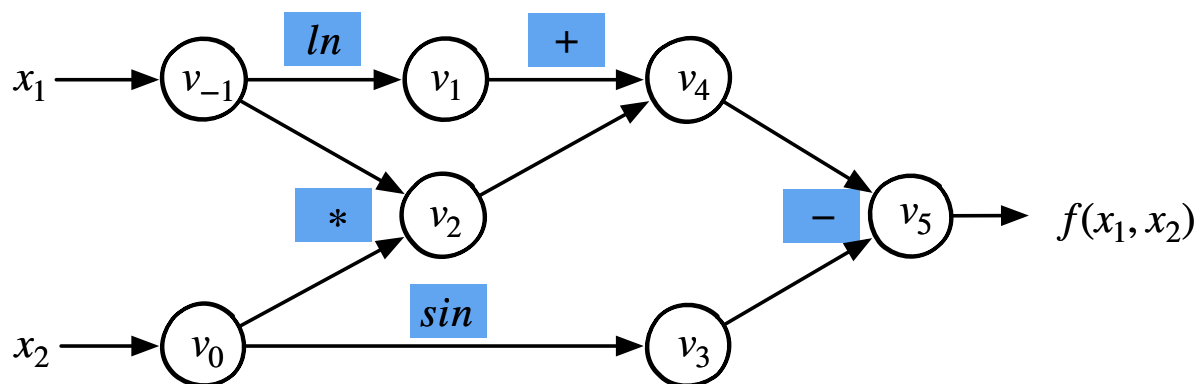
Forward Primal Trace

$v_{-1} = x_1$	$= 2$
$v_0 = x_2$	$= 5$
<hr/>	
$v_1 = \ln v_{-1}$	$= \ln 2$
$v_2 = v_{-1} \times v_0$	$= 2 \times 5$
$v_3 = \sin v_0$	$= \sin 5$
$v_4 = v_1 + v_2$	$= 0.693 + 10$
$v_5 = v_4 - v_3$	$= 10.693 + 0.959$
<hr/>	
$y = v_5$	$= 11.652$

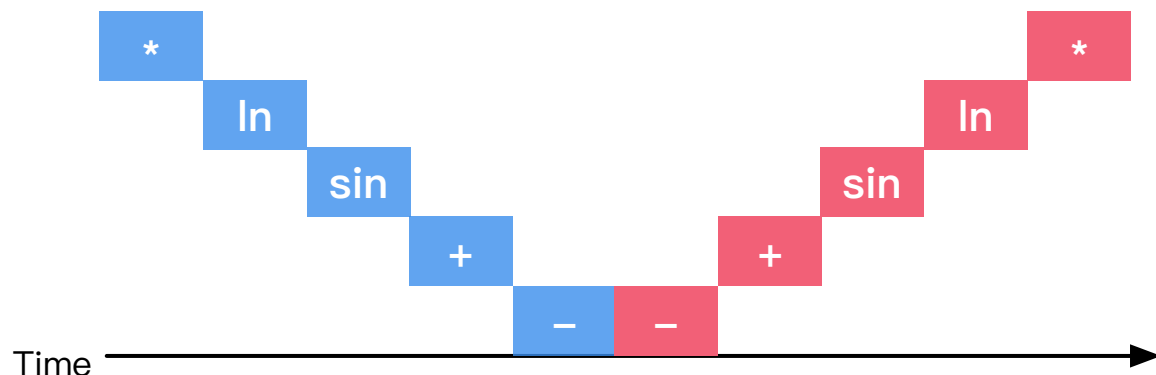
计算图的单设备算子间调度 (I)

- 计算图准确的描述了算子之间的依赖关系
- 根据计算图找到相互独立的算子进行并发调度，提高计算的并行性

$$f(x_1, x_2) = \ln(x_1) + x_1x_2 - \sin(x_2)$$

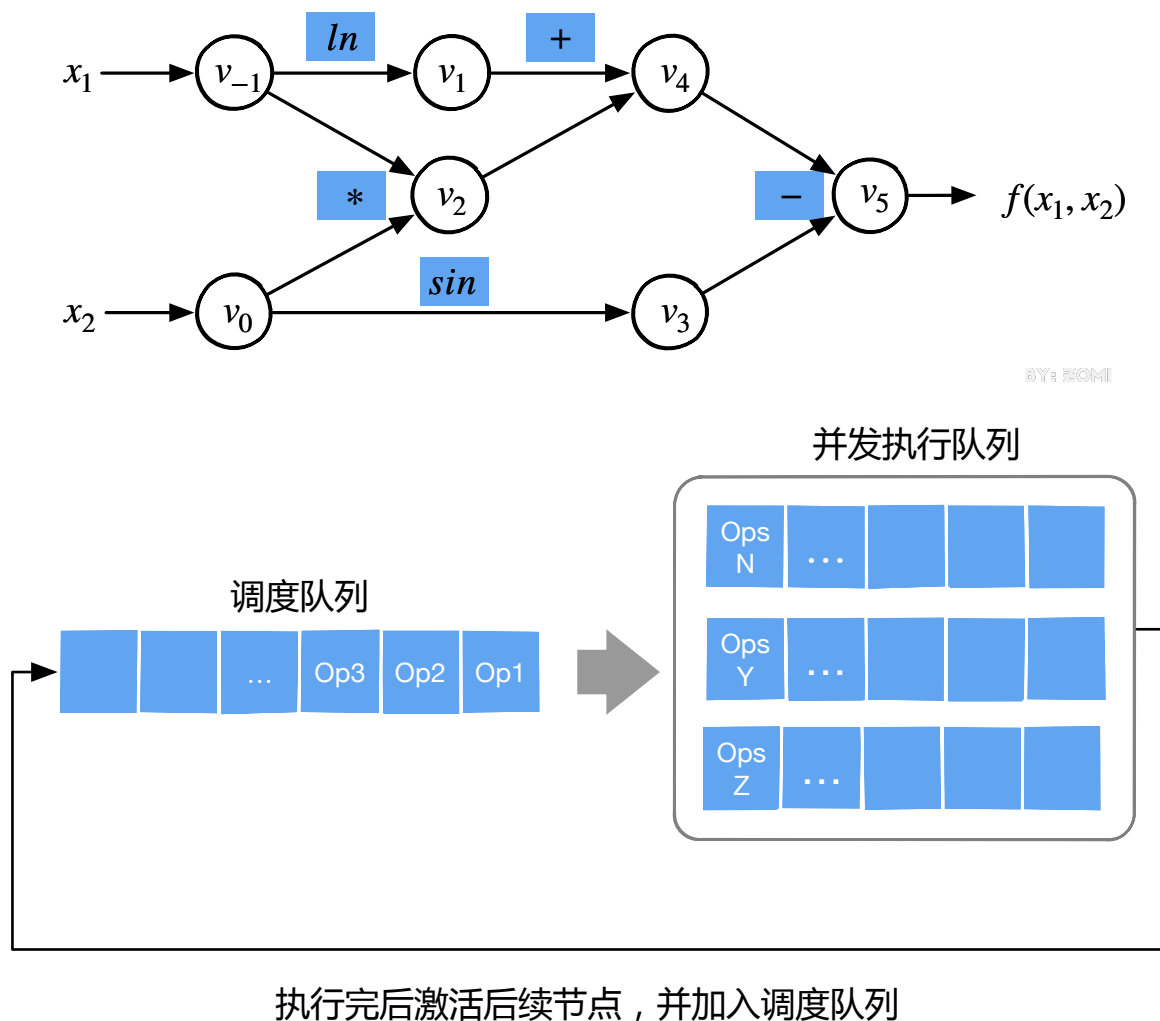


BY: ZOMI



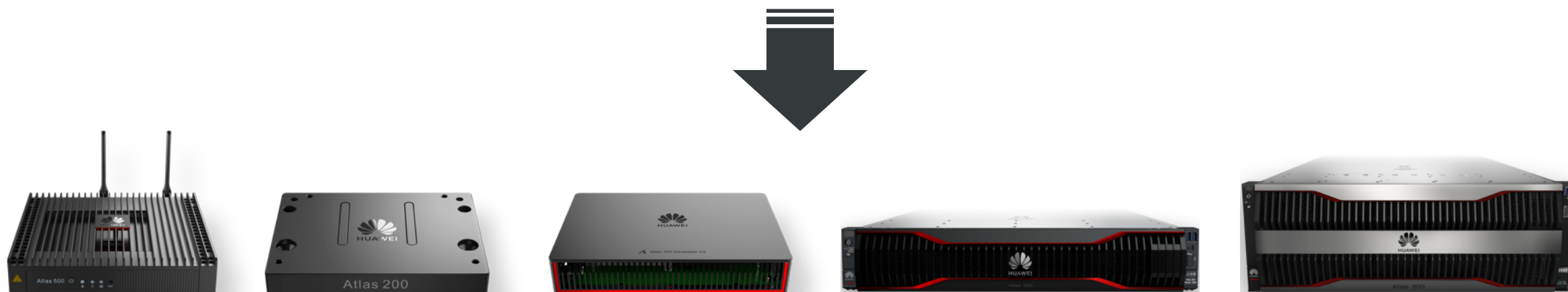
计算图的单设备算子间调度 (II)

- 数据流图准确的描述了算子之间的依赖关系
- 根据数据流图找到相互独立的算子进行并发调度，提高计算的并行性



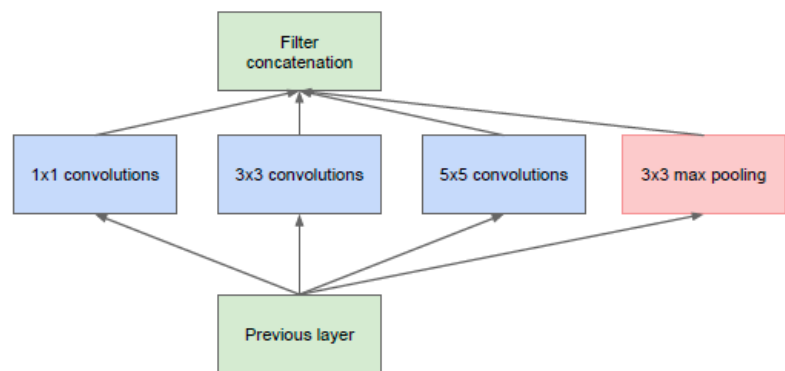
计算图的切分与多设备执行

- **计算图切分**：给定一个计算图，将计算图切分后（单算子/子图）放置到多设备上，每个设备拥有计算图的一部分。
- **插入跨设备通信**：经过切分计算图会被分成若干子图，每个子图被放置在一个设备上，进行跨设备跨子图数据传输。

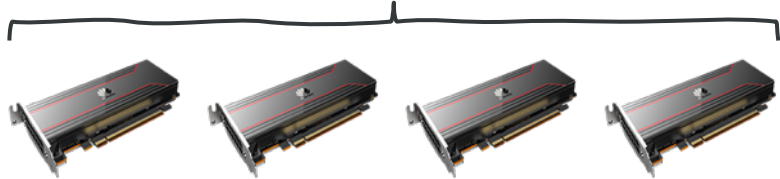


计算图的切分与多设备执行 (I)

GoogleNet Inception模块

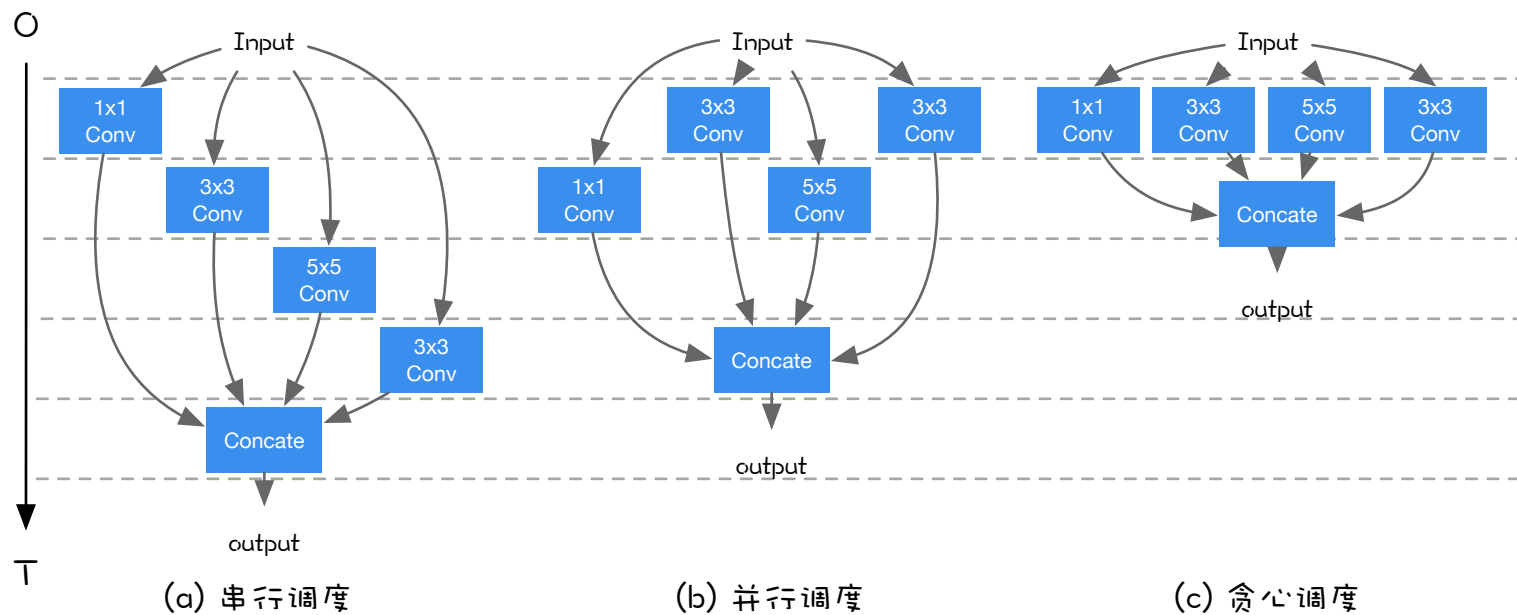


(a) Inception module, naïve version



华为昇腾 Atlas

Inception模块多设备执行策略



(a) 串行调度

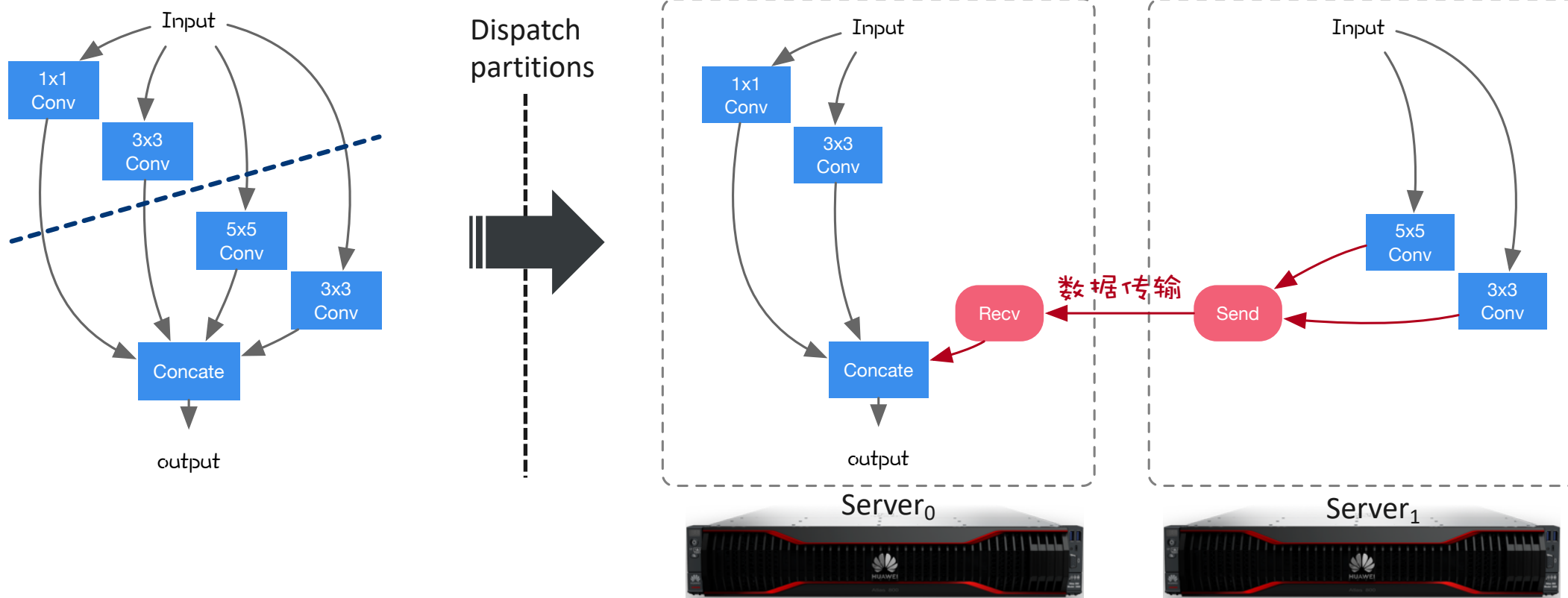
(b) 并行调度

(c) 贪心调度

多种调度模式

计算图的切分与多设备执行 (II)

跨设备的边将被自动替换成 **Send/Recv** 算子或者是通过 **通讯原语** 进行数据传输



计算图的切分与多设备执行

- **计算图切分**：给定一个计算图，将计算图切分后（单算子/子图）放置到多设备上，每个设备拥有计算图的一部分。
- **插入跨设备通信**：经过切分计算图会被分成若干子图，每个子图被放置在一个设备上，进行跨设备跨子图数据传输。



复杂组合优化问题



代价模型 (cost model)

Summary

1. 了解对于计算图进行上层的编译优化手段
2. 了解计算图最基础的调度执行模式和优化后的调度执行模式
3. 了解计算图在多设备上如何进行图切分和多设备执行



BUILDING A BETTER CONNECTED WORLD

THANK YOU

Copyright©2014 Huawei Technologies Co., Ltd. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.