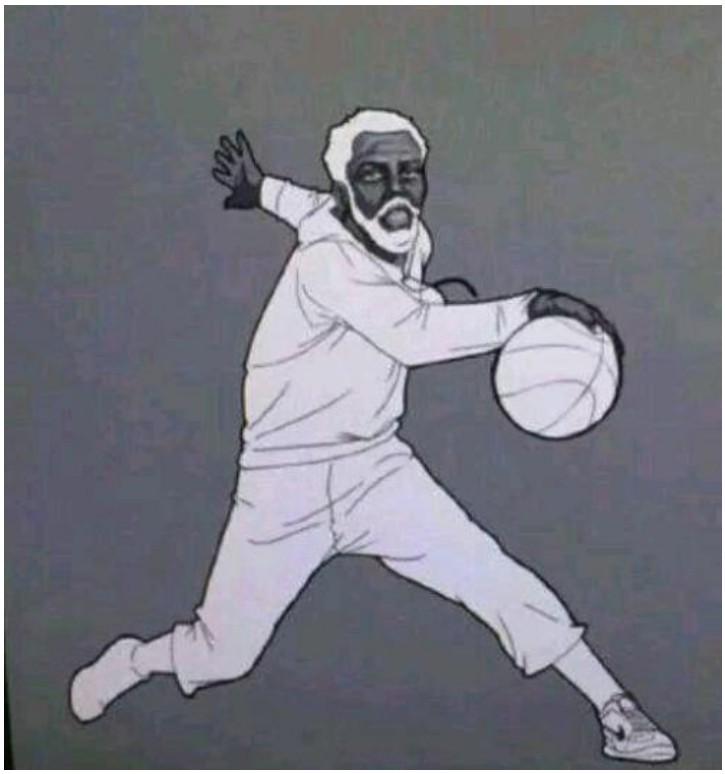
高速缓存一致性和量化分析



作者姓名: 高帅

邮箱: gshuai@mail.ustc.edu.cn

版本时间: 二〇二〇年七月二十日

写在最开始

缓存一致性 (Cache Coherent) 一直时芯片架构设计的重要问题，该文章配合仓库中的项目，来实现各种各样的 Cache 系统并比较 Cache 系统在特定场景下的性能。

本人毕业从事 AI 芯片架构设计岗位的工作，缓存一致性系统作为自己一个重要的技能栈，会广泛的学习各种发表或者开源的项目，加深自己对存储子系统的理解，服务于自己的设计工作。

目　录

第 1 章　基 础 知 识

1.1　引言

1.2　存储器层次结构

1.3　缓存优化与方法

1.3.1　1

1.3.2　2

1.3.3　3

1.4　存储区优化与技术

1.4.1　二级节标题

1. 三级节标题

（1）　四级节标题

① 五级节标题

1.5　脚注

内容[①]

[①]脚注

第 2 章　浮　动　体

2.1　三线表

三线表是《撰写手册》推荐使用的格式，如表 2.1。

表 2.1　表号和表题在表的正上方

类型	描述
挂线表	挂线表也称系统表、组织表，用于表现系统结构
无线表	无线表一般用于设备配置单、技术参数列表等
卡线表	卡线表有完全表，不完全表和三线表三种

注：表注分两种，第一种是对全表的注释，用不加阿拉伯数字排在表的下边，前面加“注：”；第二种是和表内的某处文字或数字相呼应的注，在表里面用带圈的阿拉伯数字在右上角标出，然后在表下面用同样的圈码注出来

编制表格应简单明了，表达一致，明晰易懂，表文呼应、内容一致。排版时表格字号略小，或变换字体，尽量不分页，尽量不跨节。表格太大需要转页时，需要在续表上方注明“续表”，表头页应重复排出。

2.2　插图

有的同学可能听说“LaTeX 只能使用 eps 格式的图片”，甚至把 jpg 格式转为 eps。事实上，这种做法已经过时。而且每次编译时都要要调用外部工具解析 eps，导致降低编译速度。所以我们推荐矢量图直接使用 pdf 格式，位图使用 jpeg 或 png 格式。

图 2.1　图号、图题置于图的下方

注：图注的内容不宜放到图题中。

关于图片的并排，推荐使用较新的 subcaption 宏包，不建议使用 subfigure

或 subfig 等宏包。

2.3　算法环境

模板中使用 algorithm2e 宏包实现算法环境。关于该宏包的具体用法，请阅读宏包的官方文档。

算法 2.1　算法示例 1

Data: this text
Result: how to write algorithm with LaTeX2e
initialization;
while *not at end of this document* **do**
    read current;
    if *understand* **then**
        go to next section;
        current section becomes this one;
    else
        go back to the beginning of current section;
    end
end

注意，我们可以在论文中插入算法，但是插入大段的代码是愚蠢的。然而这并不妨碍有的同学选择这么做，对于这些同学，建议用 listings 宏包。

第 3 章　数　　学

3.1　数学符号和公式

《撰写手册》要求数学符号要根据 GB/T 3102.11-1993《物理科学和技术中使用的数学符号》[①] 使用，这与 LaTeX 默认的英美国家的数学符号习惯有所差异。本模板基于 unicode-math 宏包配置数学符号，以遵循国标的规定：

1. 大写希腊字母默认为斜体，如 `\Delta`：Δ。
2. 有限增量符号 Δ（U+2206）应使用 `\increment` 命令。
3. 向量、矩阵和张量要求粗斜体，应使用 `\symbf` 命令，如 `\symbf{A}`、`\symbf{\alpha}`。
4. 数学常数和特殊函数使用正体，如圆周率 π、Γ 函数。应使用 unicode-math 宏包提供的 `\symup` 命令转为正体，如 `\symup{\pi}`。
5. 微分符号 d 使用正体，本模板提供了 `\dif` 命令。

注意，unicode-math 宏包与 amsfonts, amssymb, bm, mathrsfs, upgreek 等宏包不兼容。本模板作了处理，用户可以直接使用 `\bm`, `\mathscr`, `\upGamma`。关于数学符号更多的用法，参见 unicode-math 宏包的使用说明和符号列表 unimath-symbols。

在编辑数学公式时，最好避免直接使用字体命令，而应该定义一些语义命令取代字体命令，这样输入更简单，也让 LaTeX 代码更有可读性，而且还方便根据需要统一修改改格式。参考示例文档中的 math-commands.tex

更多的例子：

$$\mathrm{e}^{\mathrm{i}\pi} + 1 = 0 \tag{3.1}$$

$$\frac{\mathrm{d}^2 u}{\mathrm{d}t^2} = \int f(x)\,\mathrm{d}x \tag{3.2}$$

$$\arg\min_x f(x) \tag{3.3}$$

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \tag{3.4}$$

3.2　量和单位

宏包 siunitx 提供了更好的数字和单位支持：

- 12 345.678 90
- $1 \pm 2\mathrm{i}$

[①] 原 GB 3102.11-1993，根据 2017 年第 7 号公告和强制性标准整合精简结论，自 2017 年 3 月 23 日起，该标准转化为推荐性标准。

- 0.3×10^{45}
- $1.654 \times 2.34 \times 3.430$
- $\mathrm{kg} \cdot \mathrm{m} \cdot \mathrm{s}^{-1}$
- μm μm
- Ω Ω
- 10 和 20
- 10, 20 和 30
- 0.13 mm, 0.67 mm 和 0.80 mm
- 10～20
- 10 ℃～20 ℃

3.3　定理和证明

示例文件中使用 amsthm 宏包配置了定理、引理和证明等环境。用户也可以使用 ntheorem 宏包。

定义 3.1　If the integral of function f is measurable and non-negative, we define its (extended) **Lebesgue integral** by

$$\int f = \sup_{g} \int g, \tag{3.5}$$

where the supremum is taken over all measurable functions g such that $0 \leqslant g \leqslant f$, and where g is bounded and supported on a set of finite measure.

假设 3.1　The communication graph is strongly connected.

例 3.1　Simple examples of functions on $\mathbf{R}^d$ that are integrable (or non-integrable) are given by

$$f_a(x) = \begin{cases} |x|^{-a} & \text{if } |x| \leqslant 1, \\ 0 & \text{if } x > 1. \end{cases} \tag{3.6}$$

$$F_a(x) = \frac{1}{1+|x|^a}, \qquad \text{all } x \in \mathbf{R}^d. \tag{3.7}$$

Then f_a is integrable exactly when $a < d$, while F_a is integrable exactly when $a > d$.

引理 3.1 (Fatou)　Suppose $\{f_n\}$ is a sequence of measurable functions with $f_n \geqslant 0$. If $\lim_{n\to\infty} f_n(x) = f(x)$ for a.e. x, then

$$\int f \leqslant \liminf_{n\to\infty} \int f_n. \tag{3.8}$$

注　We do not exclude the cases $\int f = \infty$, or $\liminf_{n\to\infty} f_n = \infty$.

推论 3.2　Suppose f is a non-negative measurable function, and $\{f_n\}$ a sequence of non-negative measurable functions with $f_n(x) \leqslant f(x)$ and $f_n(x) \to f(x)$ for almost every x. Then

$$\lim_{n\to\infty} \int f_n = \int f. \tag{3.9}$$

命题 3.3　Suppose f is integrable on $\mathbf{R}^d$. Then for every $\epsilon > 0$:

i. There exists a set of finite measure B (a ball, for example) such that

$$\int_{B^c} |f| < \epsilon. \tag{3.10}$$

ii. There is a $\delta > 0$ such that

$$\int_E |f| < \epsilon \qquad \text{whenever } m(E) < \delta. \tag{3.11}$$

定理 3.4　Suppose $\{f_n\}$ is a sequence of measurable functions such that $f_n(x) \to f(x)$ a.e. x, as n tends to infinity. If $|f_n(x)| \leqslant g(x)$, where g is integrable, then

$$\int |f_n - f| \to 0 \qquad \text{as } n \to \infty, \tag{3.12}$$

and consequently

$$\int f_n \to \int f \qquad \text{as } n \to \infty. \tag{3.13}$$

证明　Trivial. □

Axiom of choice　Suppose E is a set and E_α is a collection of non-empty subsets of E. Then there is a function $\alpha \mapsto x_\alpha$ (a "choice function") such that

$$x_\alpha \in E_\alpha, \qquad \text{for all } \alpha. \tag{3.14}$$

Observation 1　Suppose a partially ordered set P has the property that every chain has an upper bound in P. Then the set P contains at least one maximal element.

A concise proof　Obvious. □

第4章 交 叉 问 题

4.1 数学符号和公式

第 5 章　引用文献的标注

模板使用 natbib 宏包来设置参考文献引用的格式，更多引用方法可以参考该宏包的使用说明。

5.1　顺序编码制

5.1.1　角标数字标注法

`\cite{knuth86a}` ⇒ [1]

`\citet{knuth86a}` ⇒ Knuth[1]

`\cite[42]{knuth86a}` ⇒ [1]42

`\cite{knuth86a,tlc2}` ⇒ [1-2]

`\cite{knuth86a,knuth84}` ⇒ [1,3]

5.1.2　数字标注法

`\cite{knuth86a}` ⇒ [1]

`\citet{knuth86a}` ⇒ Knuth [1]

`\cite[42]{knuth86a}` ⇒ [1]^{42}

`\cite{knuth86a,tlc2}` ⇒ [1-2]

`\cite{knuth86a,knuth84}` ⇒ [1, 3]

5.2　著者-出版年制标注法

`\cite{knuth86a}` ⇒ Knuth (1986)

`\citep{knuth86a}` ⇒ (Knuth, 1986)

`\citet[42]{knuth86a}` ⇒ Knuth (1986)^{42}

`\citep[42]{knuth86a}` ⇒ (Knuth, 1986)^{42}

`\cite{knuth86a,tlc2}` ⇒ Knuth (1986); Mittelbach et al. (2004)

`\cite{knuth86a,knuth84}` ⇒ Knuth (1986, 1984)

注意，参考文献列表中的每条文献在正文中都要被引用[4-19]。

参 考 文 献

[1] KNUTH D E. Computers and typesetting: A the TEXbook [M]. Reading, MA, USA: Addison-Wesley, 1986.

[2] MITTELBACH F, GOOSSENS M, BRAAMS J, et al. The LATEX companion [M]. 2nd ed. Reading, MA, USA: Addison-Wesley, 2004.

[3] KNUTH D E. Literate programming [J]. The Computer Journal, 1984, 27(2): 97-111.

[4] 孙立广. 顶级期刊论文摘要汇编（1999–2010）[G]. 合肥: 中国科学技术大学出版社, 2016: 222.

[5] 李泳池. 张量初步和近代连续介质力学概论 [M]. 2 版. 合肥: 中国科学技术大学出版社, 2016: 61.

[6] 刘景双. 湿地生态系统碳、氮、硫、磷生物地球化学过程 [M]. 合肥: 中国科学技术大学出版社, 2014.

[7] 程根伟. 1998 年长江洪水的成因与减灾对策 [M]//许厚泽, 赵其国. 长江流域洪涝灾害与科技对策. 北京: 科学出版社, 1999: 26-32.

[8] 陈晋镳, 张惠民, 朱士兴, 等. 蓟县震旦亚界研究 [M]//中国地质科学院天津地质矿产研究所. 中国震旦亚界. 天津: 天津科学技术出版社, 1980: 56-114.

[9] 孔庆勇, 郭红健, 孔庆和. 我国科技期刊的金字塔分层模型及发展路径初探 [J]. 中国科技期刊研究, 2015, 26(10): 1100-1103.

[10] 杨洪升. 四库馆私家抄校书考略 [J]. 文献, 2013(1): 56-75.

[11] 于潇, 刘义, 柴跃廷, 等. 互联网药品可信交易环境中主体资质审核备案模式 [J]. 清华大学学报 (自然科学版), 2012, 52(11): 1518-1521.

[12] 丁文详. 数字革命与竞争国际化 [N]. 中国青年报, 2000-11-20(15).

[13] 姜锡洲. 一种温热外敷药制备方案：中国，88105607.3 [P]. 1989-07-26.

[14] 万锦坤. 中国大学学报论文文摘（1983–1993）（英文版）[DB/CD]. 北京: 中国大百科全书出版社, 1996.

[15] 孙玉文. 汉语变调构词研究 [D]. 北京: 北京大学, 2000.

[16] 文富, 顾丽梅. 网络时代经济发展战略特征 [J]. 学术研究, 2000, 21(4): 35-40.

[17] 肖度, 等. 知识时代的企业合作经营 [M]. 北京: 北京大学出版社, 2000: 67-69.

[18] The White House. Technology for economic growth [R]. Washington, 1993.

[19] HUTSON J M. Vibrational dependence of the anisotropic intermolecular potential of argon-hydrogen chloride [J]. J. Phys. Chem., 1992, 96(11): 4237-4247.

附录 A　存储器层次结构回顾

缓存：地址离开处理器后遇到的最高一级或第一级存储器层次结构，局部性原理支撑了缓存的高性能的理论逻辑。其中**时间局部性**指当前访问的数据在不久的将来极有可能还会再次用到，**空间局部性**指当前用到的数据附近的数据（同一个数据块）也极有可能会用到。

虚拟存储器：一些数据可以存储在磁盘上，地址空间被划分为固定大小的块（页），任何时候页要么在主存储器上，要么在磁盘上。当地址页不在缓存也不再主存储器上时发生页错误，要把整个页从磁盘加载到主存储器上。页错误消耗的时间过长，处理器一般会切换任务。

A.1　缓存及缓存性能

附录 B　存储器层次结构回顾