


全面解读计算机的高级配置和电源管理原理，
使读者深入理解BIOS与Linux内核相关模块，并掌握业界先进的设计理念！

计算机系统核心开发

高级配置与电源管理最佳实践

子齐 编著

- 
- 基于ACPI 5.1规范，使用ASL描述系统配置
 - 使用系统描述表将系统配置从BIOS传递给OS
 - 使用嵌入式控制器和系统管理总线接口获取信息
 - 将系统地址映射传递给OS
 - 全局系统状态及睡眠状态之间的转换细节
 - 详解ACPI事件模型、散热区的管理及平台错误接口
 - 详述Linux内核中的ACPI实现细节

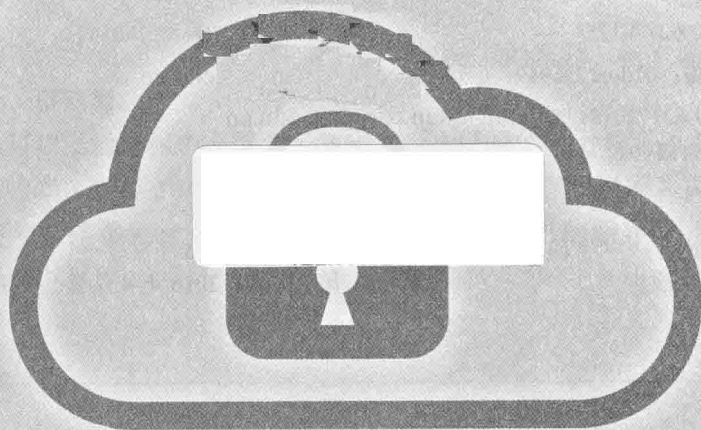
清华大学出版社



计算机系统核心开发

高级配置与电源管理最佳实践

子齐 编著



清华大学出版社

北京

内 容 简 介

本书对 UEFI 中的 ACPI 规范进行了详细介绍,是读者熟练掌握计算机系统软件 UEFI BIOS 和 Linux 内核工作原理和开发方法的必读图书之一。

本书内容包括设备及系统配置、电源和性能管理、ACPI 嵌入式控制器接口规范、ACPI 系统管理总线接口规范、处理器配置和控制、ASL 编程语言、系统描述表、ACPI 定义的设备和设备特定对象、电源和电表设备、散热管理、系统地址映射接口、睡眠和唤醒、NUMA 架构平台、ACPI 平台错误接口及 ACPI 功能实现等。

本书可作为计算机系统软件开发培训和辅导用书,也可作为大专院校师生、广大 BIOS、Linux 内核, BMC 方面的 IT 从业人员以及软、硬件开发人员的参考用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

计算机系统核心开发:高级配置与电源管理最佳实践 / 子齐编著. —北京:清华大学出版社, 2016
ISBN 978-7-302-42470-3

I. ①计… II. ①子… III. ①计算机系统—系统设计 IV. ①TP302.1

中国版本图书馆 CIP 数据核字(2015)第 310890 号

责任编辑:杨如林
装帧设计:铁海音
责任校对:胡伟民
责任印制:宋 林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京富博印刷有限公司

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:188mm×260mm 印 张:30.75 字 数:885千字

版 次:2016年4月第1版 印 次:2016年4月第1次印刷

印 数:1~3000

定 价:79.00元

产品编号:064076-01

前言

随着用户数及为了满足用户需求而开发的应用程序不断增加，包含互联网公司在内的各行各业中的企业都在不断加大对基础设施的投入。这些基础设施当中，就包含用于高速计算的服务器系统和用于保存大量数据的存储系统。如果在不影响性能的情况下能降低服务器（包含存储服务器）的整体电量消耗，一方面可以帮助企业降低运营成本，另一方面企业和个人也可以为实施国家“节能减排”战略尽一份力量。

服务器系统目前使用的UEFI BIOS和Linux内核都是非常复杂的系统软件，这两个系统软件中都包含了满足不同功能需求的各种模块。要掌握UEFI BIOS和Linux内核的工作原理和开发方法，不可能一蹴而就，只能日积月累、各个击破。当然，当企业响应国家“自主创新”战略而开发自有版权的BIOS和操作系统时，也可以参考现有的UEFI BIOS和Linux内核中的优秀设计思想和实现方法，“站在巨人的肩膀上”，“青出于蓝而胜于蓝”。

正如本书所述，ACPI模块在UEFI BIOS和Linux内核中是非常重要的一个模块，并且为BIOS和操作系统之间提供了交互接口。本书主要介绍5.1版本的ACPI规范，内容包括如何使用ASL来描述系统的配置，如何通过系统描述表将系统的配置从BIOS传递给OS，如何使用嵌入式控制器接口和系统管理总线接口来获取所需的信息，如何将系统地址映射传递给OS，以及全局系统状态及睡眠状态之间的转换细节、ACPI事件模型、散热区定义和管理、平台错误接口等，本书的最后一章介绍Linux内核中ACPI的实现细节。笔者希望本书能帮助读者更好、更快地掌握ACPI规范中的内容，并且将这些知识应用到所开发的各类软件产品中。

参与本书编写的人员还有石烁、石丽颖、石鹏泊、刘露露、何东阳、张迪、张立威、张丽丽、张劲松、蔡燕等。

在本书的编写和出版过程中，得到了清华大学出版社的大力帮助和支持，特此致谢！

本书内容比较多，涉及了BIOS、Linux内核等软件和处理器、嵌入式控制器、SMBus等硬件。书中内容与编排难免有不妥之处，殷切希望阅读本书的业界同道们能提出建设性意见，以便纠正改进。

编者

目 录

第1章 ACPI简述

1.1	首要目标	1
1.2	电源管理	2
1.3	遗留支持	3
1.4	OEM实现策略	3
1.5	电源和睡眠按钮	3
1.6	ACPI规范与结构	4
1.7	OS、平台兼容性	5
1.7.1	平台实现ACPI定义的接口	5
1.7.2	OSPM实现	5
1.7.3	OS要求	6

第2章 ACPI基本术语及概念

2.1	通用ACPI术语	7
2.2	各种状态定义	12
2.2.1	全局系统状态定义	12
2.2.2	设备电源状态定义	13
2.2.3	睡眠状态定义	14
2.2.4	处理器电源状态定义	15
2.2.5	设备和处理器性能状态定义	15
2.3	电源状态	16
2.3.1	平台电源管理特征	17
2.4	电源管理标准	18
2.5	跨设备依赖	18
2.6	硬件编程模型	18
2.6.1	固定硬件编程模型	19
2.6.2	通用硬件编程模型	19

2.7	ACPI硬件特征	20
2.8	ACPI寄存器模型	21
2.9	系统描述表头部	25

第3章 ASL编程语言

3.1	ASL语法	28
3.1.1	ASL语法表示法	28
3.2	ASL概念	29
3.2.1	ASL名字	29
3.2.2	ASL字面常量	29
3.2.3	ASL资源模板	30
3.2.4	ASL宏	31
3.2.5	ASL数据类型	31
3.2.6	ACPI名字空间	39
3.2.7	定义块	41
3.2.8	控制方法执行	43
3.2.9	资源描述符类型	46
3.3	常用ASL操作符	62
3.3.1	ACPI表管理	62
3.3.2	ASL编译控制	64
3.3.3	常量	64
3.3.4	混杂命名对象创建	65
3.3.5	操作区域和字段	71
3.3.6	字节数组字段	77
3.3.7	同步	78
3.3.8	对象参考	79
3.3.9	整数运算	79
3.3.10	逻辑操作符	82
3.3.11	方法执行控制	83
3.3.12	数据类型转换和处理	87
3.3.13	资源描述符宏	93
3.3.14	控制方法对象	108
3.4	预定义的对象	109
3.4.1	_GL (Global Lock Mutex)	109
3.4.2	_OSI (Operating System Interface)	109
3.4.3	_OS (OS Name Object)	112

3.4.4 _REV (Revision Data Object)	112
3.5 转化ASL到AML	112

第4章 系统描述表架构

4.1 概要	115
4.2 ACPI系统描述表	117
4.2.1 保留位和字段	117
4.2.2 兼容性	118
4.2.3 地址格式	118
4.2.4 UUID (Universal Uniform Identifiers)	119
4.2.5 RSDP	119
4.2.6 RSDT	120
4.2.7 XSDT	121
4.2.8 FADT	121
4.2.9 FACS	140
4.2.10 DSDT	143
4.2.11 SSDT	143
4.2.12 MADT	144
4.2.13 SBST	151
4.2.14 ECDT	151
4.2.15 SRAT	153
4.2.16 SLIT	155
4.3 其他ACPI表	156
4.3.1 MCFG	156
4.3.2 HPET	157
4.3.3 DMAR	158
4.3.4 微软软件授权表	158
4.3.5 SPMI	159
4.3.6 SPCR	160

第5章 ACPI事件编程模型

5.1 ACPI事件编程模型组件	165
5.2 ACPI事件类型	166
5.3 固定事件处理	166
5.4 通用目的事件处理	167

5.4.1	针对GPE处理的 _Exx、_Lxx和_Qxx方法	170
5.4.2	GPE唤醒事件	171
5.5	设备对象通知	173

第6章 设备及系统配置

6.1	设备标识对象	176
6.1.1	_ADR (Address)	177
6.1.2	_CID (Compatible ID)	178
6.1.3	_CLS (Class Code)	179
6.1.4	_DDN (DOS Device Name)	179
6.1.5	_HID (Hardware ID)	179
6.1.6	_HRV (Hardware Revision)	180
6.1.7	_MLS (Multiple Language String)	180
6.1.8	_SUB	181
6.1.9	_STR (String)	181
6.1.10	_UID (Unique ID)	182
6.1.11	_SUN (Slot User Number)	182
6.2	设备配置对象	183
6.2.1	_CDM (Clock Domain)	184
6.2.2	_CRS (Current Resource Settings)	184
6.2.3	_DIS (Disable)	185
6.2.4	_DMA (Direct Memory Access)	185
6.2.5	_DSD (Device Specific Data)	187
6.2.6	_FIX (Fixed Register Resource Provider)	187
6.2.7	_GSB (Global System Interrupt Base)	189
6.2.8	_HPP (Hot Plug Parameters)	190
6.2.9	_HPX (Hot Plug Parameters Extensions)	192
6.2.10	_MAT (Multiple APIC Table Entry)	195
6.2.11	_OSC (Operating System Capabilities)	196
6.2.12	_PRS (Possible Resource Setting)	202
6.2.13	_PRT (PCI Routing Table)	202
6.2.14	_PXM (Proximity)	205
6.2.15	_SLI (System Locality Information)	205
6.2.16	_CCA (Cache Coherency Attribute)	207
6.2.17	_SRS (Set Resource Settings)	208
6.3	设备插入、移走和状态对象	209

6.3.1	_EDL(Eject Device List)	210
6.3.2	_EJD(Eject Dependent Device)	211
6.3.3	_EJx(Eject)	212
6.3.4	_LCK(Lock)	212
6.3.5	_OST(OSPM Status Indication)	212
6.3.6	_RMV(Remove)	216
6.3.7	_STA (Status)	216
6.4	其他对象和控制方法	217
6.4.1	_INI (Init)	218
6.4.2	_REG (Region)	218
6.4.3	_BBN (Base Bus Number)	220
6.4.4	_SEG (Segment)	220
6.4.5	_GLK (Global Lock)	222
6.5	系统配置对象	222
6.6	ISA设备配置	222

第7章 电源和性能管理

7.1	声明电源资源对象	228
7.2	为电源资源定义的子对象	228
7.2.1	_OFF	229
7.2.2	_ON	229
7.2.3	_STA (Status)	229
7.3	设备电源管理对象	229
7.3.1	_DSW (Device Sleep Wake)	231
7.3.2	_PSx (Power State x)	231
7.3.3	_PSC (Power State Current)	231
7.3.4	_PR0 (Power Resource for D0)	232
7.3.5	_PR1 (Power Resource for D1)	232
7.3.6	_PR2 (Power Resource for D2)	233
7.3.7	_PR3 (Power Resource for D3 _{hot})	233
7.3.8	_PRW (Power Resource for Wake)	233
7.3.9	_PSW (Power State Wake)	234
7.3.10	_IRC (In Rush Current)	235
7.3.11	_S1D (S1 Device State)	235
7.3.12	_S2D (S2 Device State)	235
7.3.13	_S3D (S3 Device State)	236

7.3.14	_S4D (S4 Device State)	236
7.3.15	_S0W (S0 Device Wake State)	237
7.3.16	_S1W (S1 Device Wake State)	237
7.3.17	_S2W (S2 Device Wake State)	237
7.3.18	_S3W (S3 Device Wake State)	238
7.3.19	_S4W (S4 Device Wake State)	238
7.4	OEM提供的系统级别控制方法	238
7.4.1	_PTS	239
7.4.2	系统_Sx状态	239
7.4.3	_SWS (System Wake Source)	242
7.4.4	_TTS (Transition To State)	242
7.4.5	_WAK (System Wake)	243
7.5	OSPM使用_PTS、_TTS和_WAK	243
7.6	控制设备电源	244
7.6.1	获取设备电源能力	244
7.6.2	设置设备电源状态	244
7.6.3	获取设备电源状态	245
7.6.4	唤醒计算机	245
7.6.5	调制解调器设备电源管理示例	245

第8章 ACPI嵌入式控制器接口规范

8.1	嵌入式控制器接口描述	250
8.2	嵌入式控制器寄存器描述	251
8.2.1	嵌入式控制器状态寄存器	251
8.2.2	嵌入式控制器命令寄存器	252
8.2.3	嵌入式控制器数据寄存器	252
8.3	嵌入式控制器命令集	252
8.3.1	读嵌入式控制器	252
8.3.2	写嵌入式控制器	252
8.3.3	突发使能嵌入式控制器	253
8.3.4	突发屏蔽嵌入式控制器	253
8.3.5	查询嵌入式控制器	253
8.4	中断模型	254
8.4.1	事件中断模型	254
8.4.2	命令中断模型	254
8.5	嵌入式控制器接口算法	255

8.6	嵌入式控制器描述信息	255
8.7	通过嵌入式控制器实现的SMBus主控制器接口	255
8.7.1	寄存器描述	256
8.7.2	协议描述	259
8.8	SMBus设备	261
8.8.1	SMBus设备访问限制	262
8.8.2	SMBus设备命令访问限制	262
8.9	在ACPI名字空间中定义一个嵌入式控制器设备	262
8.10	在ACPI名字空间中定义一个EC SMBus主控制器	266

第9章 ACPI系统管理总线接口规范

9.1	SMBus概要	268
9.1.1	SMBus从地址	268
9.1.2	SMBus协议	269
9.1.3	SMBus状态码	269
9.1.4	SMBus命令值	269
9.2	ASL代码访问SMBus	270
9.2.1	声明SMBus主控制器对象	270
9.2.2	声明SMBus设备	271
9.2.3	声明SMBus操作区域	271
9.2.4	声明SMBus字段	272
9.2.5	声明和使用SMBus数据缓冲区	274
9.3	使用SMBus协议	275
9.3.1	快速读/写 (SMBQuick)	275
9.3.2	发送/接收字节 (SMBSendReceive)	276
9.3.3	读/写字节 (SMBByte)	277
9.3.4	读/写双字节 (SMBWord)	278
9.3.5	读/写块 (SMBBlock)	278
9.3.6	双字节过程调用 (SMBProcessCall)	279
9.3.7	块过程调用 (SMBBlockProcessCall)	280

第10章 处理器配置和控制

10.1	处理器电源状态	282
10.1.1	处理器电源状态C0	283
10.1.2	处理器电源状态C1	285

10.1.3	处理器电源状态C2	285
10.1.4	处理器电源状态C3	285
10.1.5	额外处理器电源状态	286
10.2	刷新缓存	287
10.3	声明处理器	287
10.3.1	_PDC (Processor Driver Capabilities)	288
10.3.2	处理器电源状态控制	289
10.3.3	处理器风门控制	292
10.3.4	处理器性能控制	297
10.4	处理器聚合设备	302
10.4.1	逻辑处理器空载	302
10.4.2	_PUR (Processor Utilization Request)	302

第11章 ACPI定义的设备和设备特定对象

11.1	_SI系统指示标记	304
11.1.1	_SST (System Status)	305
11.1.2	_MSG (Message)	305
11.1.3	_BLT (Battery Level Threshold)	305
11.1.4	ASL代码示例	305
11.2	电池设备	306
11.3	控制型外盖设备	306
11.3.1	_LID	307
11.3.2	实现示例	307
11.4	控制型电源和睡眠按钮设备	310
11.5	嵌入式控制器设备	312
11.6	通用容器设备	313
11.7	ATA控制器设备	313
11.7.1	_GTF (Get Task File)	313
11.7.2	IDE控制器设备	314
11.7.3	SATA控制器设备	317
11.8	GPE块设备	318
11.9	模块设备	319
11.10	内存设备	322
11.10.1	地址解码	323
11.10.2	内存设备的_OSC定义	323
11.11	_UPC (USB Port Capabilities)	324

11.12	USB 2.0主控制器和 _UPC、_PLD	326
11.13	设备对象名冲突	327
11.13.1	_DSM (Device Specific Method)	327
11.14	PC/AT RTC/CMOS设备	327
11.14.1	PC/AT兼容RTC/CMOS设备 (PNP0B00)	327
11.14.2	Intel PIIX4兼容RTC/CMOS设备 (PNP0B01)	328
11.14.3	Dallas Semiconductor兼容RTC/CMOS设备 (PNP0B02)	329
11.15	I/O APIC设备	329

第12章 电源和电表设备

12.1	电池管理	330
12.1.1	电池通讯	331
12.1.2	电池容量	331
12.1.3	电池容量规格	331
12.1.4	低电量容量级别	332
12.1.5	电池校准	333
12.2	智能电池子系统	334
12.2.1	ACPI智能电池状态改变通知要求	335
12.2.2	智能电池对象	336
12.2.3	_SBS(Smart Battery Subsystem)	336
12.3	控制型电池	338
12.3.1	电池事件	339
12.3.2	电池控制方法	339
12.3.3	ASL代码示例	347
12.4	AC适配器和电源对象	348
12.4.1	_PSR(Power Source)	349
12.4.2	_PCL(Power Consume List)	349
12.4.3	_PIF(Power Source Information)	350
12.4.4	_PRL(Power Source Redundancy List)	350
12.5	电表	350
12.5.1	_PMC(Power Meter Capabilities)	351
12.5.2	_PTP(Power Trip Points)	352
12.5.3	_PMM(Power Meter Measurement)	352
12.5.4	_PAI(Power Averaging Interval)	352
12.5.5	_GAI(Get Averaging Interval)	353
12.5.6	_SHL(Set Hardware Limit)	353

12.5.7	_GHL(Get Hardware Limit)	353
12.5.8	_PMD(Power Metered Devices)	353
12.6	电源和电表名字空间示例	353

第13章 散热管理

13.1	散热控制	355
13.1.1	主动和被动冷却模型	356
13.1.2	性能和节省电量	356
13.1.3	噪声	357
13.1.4	多个散热区	357
13.1.5	主动、被动和临界策略	357
13.1.6	动态地改变冷却温度触发点	358
13.1.7	识别温度改变	359
13.1.8	主动冷却	360
13.1.9	被动冷却	361
13.1.10	临界关机	362
13.2	冷却首选项	362
13.2.1	评估散热设备列表	364
13.2.2	评估设备散热关系信息	364
13.2.3	风扇设备通知	364
13.3	风扇设备	364
13.3.1	_FIF(Fan Information)	365
13.3.2	_FPS(Fan Performance States)	366
13.3.3	_FSL(Fan Set Level)	366
13.3.4	_FST(Fan Status)	366
13.4	散热对象	367
13.4.1	_ACx(Active Cooling)	367
13.4.2	_ALx(Active List)	368
13.4.3	_ART(Active Cooling Relationship Table)	368
13.4.4	_CRT(Critical Temperature)	370
13.4.5	_DTI(Device Temperature Indication)	370
13.4.6	_HOT(Hot Temperature)	370
13.4.7	_NTT(Notification Temperature Threshold)	370
13.4.8	_PSL(Passive List)	370
13.4.9	_PSV(Passive)	371
13.4.10	_RTV(Relative Temperature Values)	371

13.4.11	_SCP(Set Cooling Policy).....	371
13.4.12	_TC1(Thermal Constant 1).....	374
13.4.13	_TC2(Thermal Constant 2).....	374
13.4.14	_TMP(Temperature).....	374
13.4.15	_TPT(Trip Point Temperature).....	375
13.4.16	_TRT(Thermal Relationship Table).....	375
13.4.17	_TSP(Thermal Sampling Period).....	375
13.4.18	_TST(Temperature Sensor Threshold).....	375
13.4.19	_TZD(Thermal Zone Devices).....	376
13.4.20	_TZM(Thermal Zone Member).....	376
13.4.21	_TZP(Thermal Zone Polling).....	376
13.5	OS设备驱动散热接口.....	376
13.6	散热区接口要求.....	377
13.7	散热区示例.....	377
13.7.1	基础散热区.....	378
13.7.2	多速率风扇.....	379
13.7.3	具有多个设备的散热区.....	381

第14章 系统地址映射接口

14.1	INT 15h, E820h——查询系统地址映射.....	387
14.2	E820假设和限制.....	389
14.3	UEFI启动服务功能.....	389
14.4	UEFI假设和限制.....	390
14.5	地址映射示例.....	390
14.6	操作系统调用示例.....	391

第15章 唤醒和睡眠

15.1	睡眠状态.....	393
15.1.1	S1睡眠状态.....	395
15.1.2	S2睡眠状态.....	396
15.1.3	S3睡眠状态.....	396
15.1.4	S4睡眠状态.....	397
15.1.5	S5软关机状态.....	398
15.1.6	从工作状态转换到睡眠状态.....	398
15.1.7	从工作状态转换到软关机状态.....	399

15.2	刷新缓存	399
15.3	初始化	400
15.3.1	将系统置于ACPI模式	401
15.3.2	BIOS对内存进行初始化	402
15.3.3	加载OS	404
15.3.4	退出ACPI模式	405

第16章 NUMA架构平台

16.1	NUMA节点	406
16.2	系统位置	407
16.2.1	SRAT定义	407
16.2.2	系统资源关联更新	407
16.3	系统位置距离信息	407
16.3.1	在线插拔	408
16.3.2	对已存在位置的影响	408

第17章 ACPI平台错误接口 (APEI)

17.1	硬件错误和错误源	409
17.2	OSPM和系统固件之间的关系	410
17.3	错误源发现	410
17.3.1	启动错误源	410
17.3.2	ACPI错误源	411
17.4	固件首先处理错误	420
17.5	错误序列化	421
17.5.1	序列化动作表	421
17.5.2	操作	425
17.6	错误注入	429
17.6.1	EINJ	429
17.6.2	注入指令项	430
17.6.3	注入指令	431
17.6.4	触发动作表	433
17.6.5	错误注入操作	433

第18章 ACPI功能实现

18.1	架构概要	436
18.1.1	ACPICA系统概要	436
18.1.2	ACPICA子系统架构	438
18.1.3	OS服务层架构	440
18.2	设计细节	441
18.2.1	ACPI名字空间基本原理	441
18.2.2	执行模型	443
18.2.3	原理和理论	449
18.3	实现细节	450
18.3.1	主机OS初始化顺序要求	450
18.3.2	ACPICA初始化顺序要求	451
18.3.3	多线程支持	456
18.3.4	通用目的事件支持	461
18.3.5	混杂ACPICA行为	463
18.4	ACPICA部署向导	464
18.4.1	使用ACPICA子系统接口	464
18.4.2	实现OS服务层	469