# 嵌入式Linux系统中的电源管理研究

## 摘要

[嵌入式Linux系统中的电源管理研究 1](#__RefHeading___Toc403680798)

[摘要 1](#__RefHeading___Toc403680799)

[第一章 绪论 2](#__RefHeading___Toc403680800)

[1.1 研究背景 2](#__RefHeading___Toc403680801)

[1.2 研究现状 2](#__RefHeading___Toc403680802)

[1.3 本文研究工作 2](#__RefHeading___Toc403680803)

[1.4 论文结构 2](#__RefHeading___Toc403680804)

[第二章 嵌入式Linux系统电源管理技术分析 2](#__RefHeading___Toc403680805)

[2.1 文献综述 2](#__RefHeading___Toc403680806)

[2.2 存在问题 2](#__RefHeading___Toc403680807)

[2.3 研究路线 2](#__RefHeading___Toc403680808)

[2.4 相关理论与技术简介 2](#__RefHeading___Toc403680809)

[第三章 嵌入式Linux系统电源管理的研究与设计 2](#__RefHeading___Toc403680810)

[3.1 嵌入式Linux系统电源管理架构 2](#__RefHeading___Toc403680811)

[3.2 设备驱动的Runtime PM设计 3](#__RefHeading___Toc403680812)

[3.2.1 通用设计思想 3](#__RefHeading___Toc403680813)

[3.2.2 Atmel MXT1664触摸屏控制器驱动的Runtime PM设计 6](#__RefHeading___Toc403680814)

[**3.2.2.1** **MXT1664的硬件特性** 6](#__RefHeading___Toc403680815)

[**3.2.2.2** **软件设计目标** 7](#__RefHeading___Toc403680816)

[**3.2.2.3** **驱动流程** 7](#__RefHeading___Toc403680817)

[第四章 嵌入式Linux系统电源管理的实现 9](#__RefHeading___Toc403680818)

[第五章 总结与展望 9](#__RefHeading___Toc403680819)

## 第一章 绪论

### 1.1 研究背景

### 1.2 研究现状

### 1.3 本文研究工作

### 1.4 论文结构

## 第二章 嵌入式Linux系统电源管理技术分析

### 2.1 文献综述

### 2.2 存在问题

### 2.3 研究路线

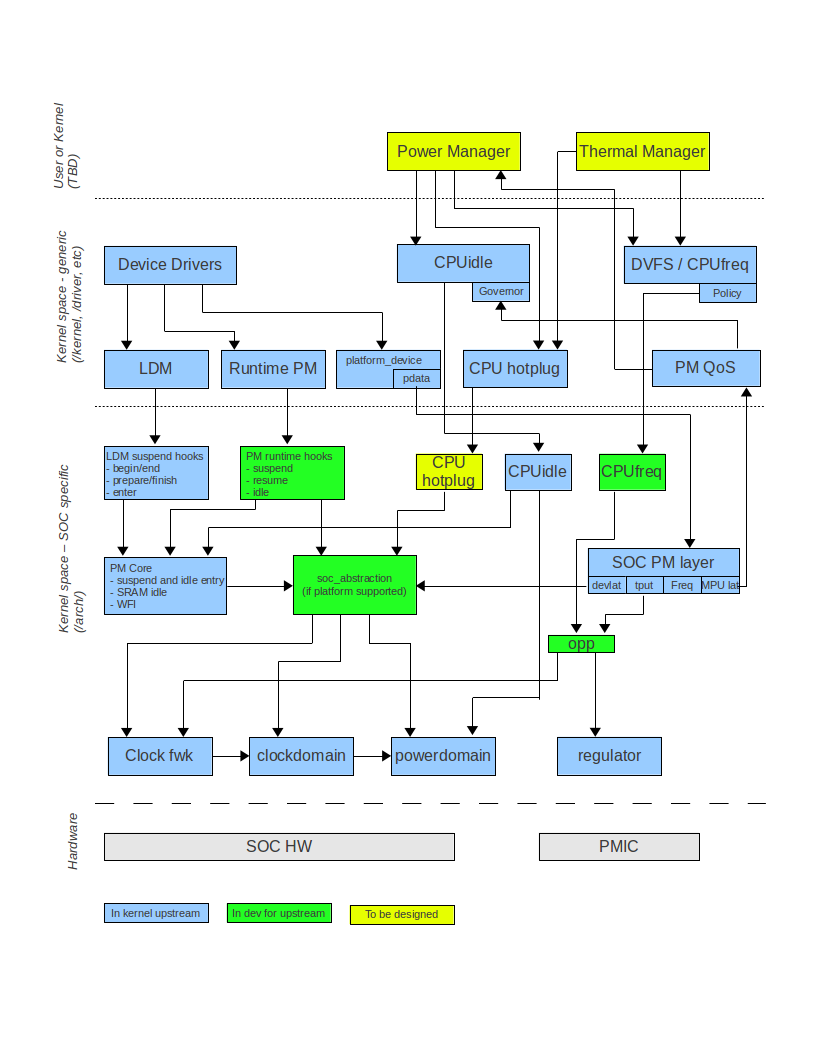
### 2.4 相关理论与技术简介

## 第三章 嵌入式Linux系统电源管理的研究与设计

### 3.1嵌入式Linux系统电源管理架构

#### 3.1.1 电源管理架构概述

嵌入式Linux系统的电源管理架构是在通用的Linux内核电源管理基础之上加入了用户空间电源管理机制而成的。整体架构可以用下图表示。其中，从下往上依次为硬件层，Linux内核层和用户空间层。而Linux内核层又可以根据与硬件的依赖关系分为硬件SoC平台相关和平台通用两层。



硬件层通常指的是作为嵌入式Linux系统硬件核心的SoC硬件以及与之相配套的电源管理芯片PMIC。SoC通常在硬件上提供了一系列电源管理特性比如处理器核的低功耗状态、处理器核心和外设总线时钟频率的设置寄存器、门控时钟（clock gating）等。PMIC通常由SoC硬件以及Linux内核中的PMIC驱动协调控制实现系统中各组件供电的打开与关闭以及电压的调整。由此可以实现处理器核心在低功耗状态下相应的电压调整。除此之外，在现代SoC集成度越来越高的趋势下，越来越多的处理器核如无线通信基带处理器，数字信号处理器，WiFi／蓝牙／GPS通信处理器等都集成到了同一个SoC系统内。因此SoC内也出现了专门负责资源管理到处理器核心来统一调度SoC内不同处理器核的电源管理状态。在HP Pro Slate 8项目中，所选取的SoC平台高通APQ8074AB中就包含了这样的一颗处理器核，称为Resource and Powere Manager (RPM)核。

［APQ8074AB Power Managment HW block diagram］

从图中可以看到，PM8941 PMIC和APQ8074AB通过System Power Management Interface(SPMI)接口连接，通过PMIC的时钟和电源电路提供了SoC工作所需的时钟信号和处理器核心以及片上集成的DDR内存控制器、外部eMMC控制器、I2C总线控制器等主要组件的电源输入。当APQ8074AB里的Krate应用处理器核心的工作频率在正常工作范围内动态变化时，对应的PM8941里的Power Regulaor会根据驱动预先的设定值相应调整处理器核心的电压以达到功率的控制；当SoC除RPM以外的处理器核们进入深度休眠状态时，大部分的电源和系统主要的19.2MHz的晶振输入信号将被关闭，只保留RPM用以唤醒处理器核的电源和一个低频率的32.768K的时钟信号以达到整个SoC的最低功耗。除此之外，APQ8074AB上的时钟控制寄存器和PM8941的电源电路控制寄存器可用于在系统运行时由Linux内核的电源管理框架来实现设备的时钟和电源的控制，这些是实现运行时电源管理(Runtime PM)和系统休眠和唤醒的核心电源管理(Linux PM Core)的基础。APQ8074AB还提供了如WFI(Wait for Interrupt)这样的低功耗状态用以支持Linux内核的CPUIdle机制，实现处理器空闲情况下的低功耗管理。

在硬件层之上的就是嵌入式Linux系统的内核。在目前最新的3.x版本上它包含了实现电源管理所需要的时钟框架和电源驱动框架(Linux voltage and current regulator framework)。在此基础之上，它提供了系统休眠与唤醒机制，即Linux核心电源管理、基于Runtime PM框架的设备动态电源管理和一系列针对处理器核心的动态电源管理机制，如CPUIdle，CPUFreq和CPU Hotplug。Linux内核中的电源管理架构设计遵从Linux内核的一般设计准则，即分为上层硬件平台无关和底层硬件平台相关这两层。通常硬件平台无关层通过sysfs这样的接口向用户空间提供控制接口，使得用户空间的应用程序可以控制系统的电源管理行为。而硬件平台相关层则由系统SoC硬件平台代码实现，用以向各个电源管理模块的硬件平台无关层提供控制特定SoC硬件平台的接口，如调节处理器核的运行频率和电压。本节下面的内容将以APQ8074AB硬件平台为例介绍Linu核心电源管理、CPUIdle、CPUFreq和CPU Hotplug模块的设计。至于Runtime PM的设计，将在3.2节中详细介绍。

在Linux内核之上运行的是用户空间的电源管理模块，通常可以分为两类：一是负责控制系统电源状态的电源管理模块，比如控制系统的运行和休眠状态切换；二是负责调节系统的散热性能的热能管理模块，比如在系统温度过高时降低处理器核的运行频率和电压以达到散热的目的。

#### 3.1.2 Linux核心电源管理

Linux核心电源管理主要负责控制内核层面上的系统睡眠与唤醒[1]。这里的睡眠指系统从正常工作状态切换到某一低功耗状态。在3.x版本到Linux内核中，定义了三种低功耗状态：待机（Standby）、睡眠（Suspend）和休眠（Hibernate）[2]。其中，待机状态下处理器核处于停止状态而大部分硬件设备也处于空闲状态。这种状态带来的功耗降低效果最少但回到正常工作状态但延迟最短，通常在1秒内。睡眠状态也被称为Suspend-to-RAM。在这种状态下，大部分硬件设备都处于电源关闭或者低功耗状态，处理器核的电源通常是关闭的。程序上下文保存在DRAM中，DRAM处于低功耗的自刷新状态。整个系统只保留用以唤醒系统的模块的供电。这些模块将用以在特定条件下，比如电源键被按下，唤醒处理器核并从DRAM中恢复程序上下文。与待机状态相比，睡眠状态带来的功耗降低效果大很多，通常对于嵌入式Linux系统来说，整个系统的耗电在10mA以下。从睡眠状态恢复到正常状态的延迟与待机状态相比略微增加，但仍可以保持在1秒内。休眠模式通常被称为Suspend-to-Disk。在该状态下，程序上下文保存在外部非易失性存储设备中，比如硬盘。因此，整个系统可以完全关闭电源从而达到最低功耗状态。但同时返回正常工作状态但延迟也最长，可以达到睡眠状态的数十倍。对于嵌入式设备，尤其是手持智能终端设备来说，时刻保持对用户对高度响应至关重要，因此，在这类系统里，Linux核心电源管理主要负责正常工作状态和睡眠状态之间的切换。

#### 3.1.3 CPUIdle模块设计

#### 3.1.4 CPUFreq模块设计

#### 3.1.5 CPU Hotplug模块设计

### 3.2设备驱动的Runtime PM设计

#### 3.2.1通用设计思想

Runtime PM框架是按照Linux内核中的设备模型设计的。在该设备模型中，设备之间是一个抽象的总线树形的关系：设备挂载在某一类型的总线上，它们是总线控制器的子设备，而它们本身又是下一级总线设备的父设备。例如在下图中，CPU下由PCI总线控制器，Display总线控制器和I2C总线控制器3个子设备。而在PCI总线控制器下又有3个PCI设备作为它的子设备。其中USB控制器又是下一级USB总线上各个USB设备的父设备。

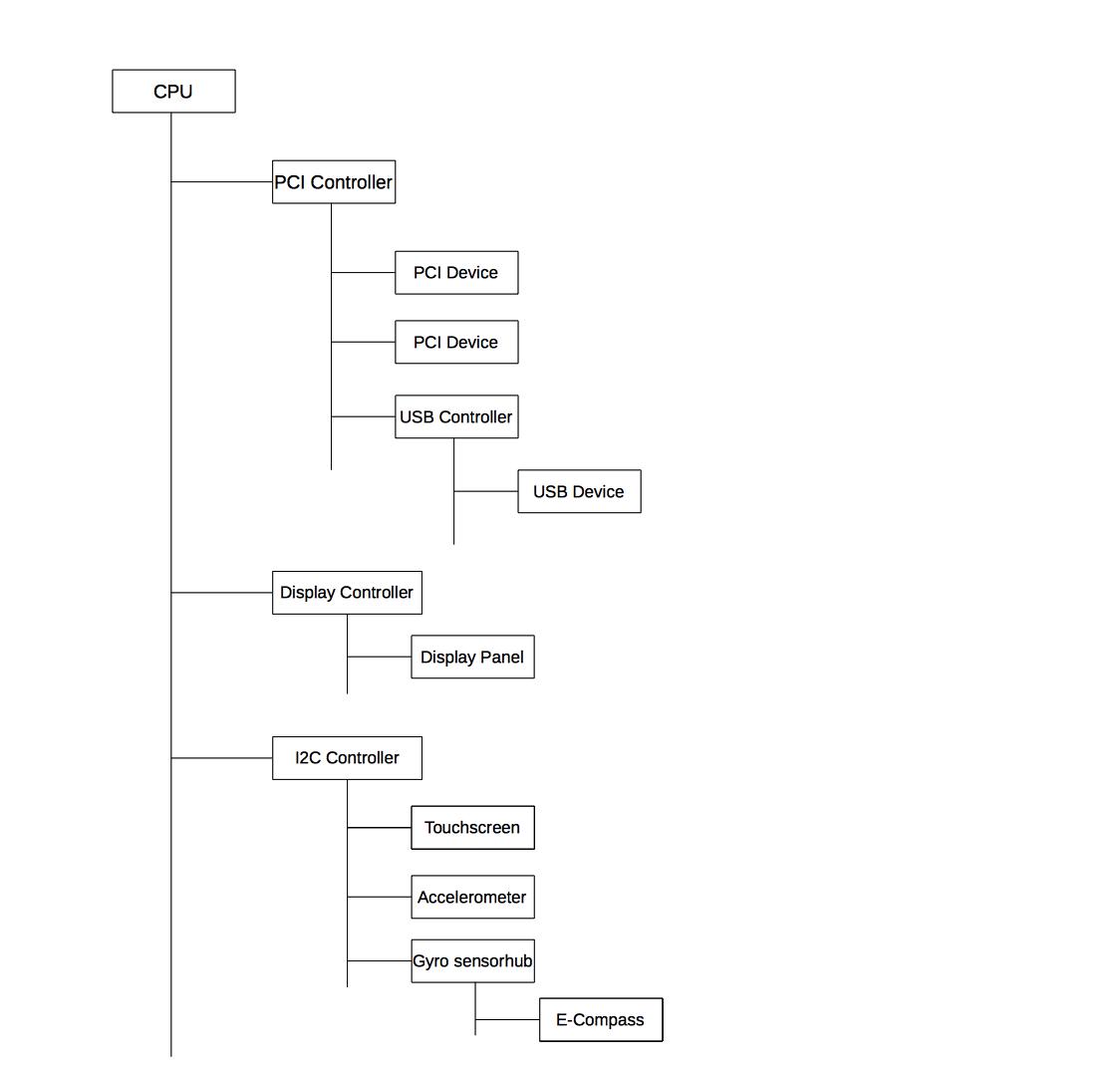


图 3‑2 Linux内核设备模型

Figure 3‑2 Linux device model

基于这一模型，Runtime PM为每一个设备定义了两个计数器：usage\_count和child\_count。其中，usage\_count用来纪录当前设备的使用量计数：每当当前设备的使用量发生改变，如开始使用或结束使用，usage\_count相应地增加或减少1；child\_count用来纪录当前设备的活跃(active)子设备的计数。

在Runtime PM框架中，设备有3个电源管理状态：活跃(active)、空闲(idle)和休眠(suspended)。Runtime PM在初始化时认为每一个设备都是处于suspended状态。当设备被用户开始使用后，设备的状态由suspended转换为active，同时相应地增加usage\_count和当前设备的父设备的child\_count。如果父设备此时处于suspended或idle状态，父设备也将转换为active状态，依此类推。当设备的所有用户使用结束，usage\_count减少到0并且同时child\_count也是0的话，设备将由active状态转换到idle状态。如果在idle状态下设备可以休眠，Runtime PM将会把设备切换到suspended状态。如果此时又出现用户使用当前设备或出现active状态的子设备，当前设备将切换到active状态。Usage\_count的增加和减少是通过调用pm\_runtime\_get()和pm\_runtime\_put()以及这两个函数的变种函数实现的。整个状态切换过程如下图所示：

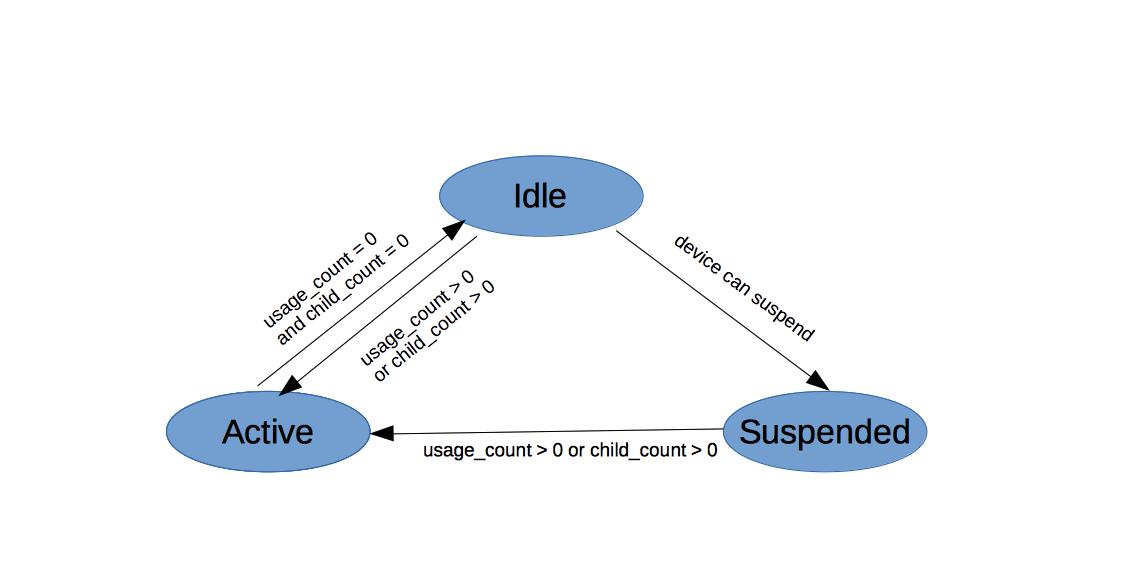


图 3‑2 Runtime PM状态

Figure 3‑2 Runtime PM states

在上述状态述切换过程中，设备驱动可以实现三个相应的回调函数(callback function)：runtime\_suepnd()，runtime\_resume()和 runtime\_idle()。函数的原型声明如下：

struct dev\_pm\_ops {

...

int (\*runtime\_suspend)(struct device \*dev);

int (\*runtime\_resume)(struct device \*dev);

int (\*runtime\_idle)(struct device \*dev);

...

};

其中，runtime\_suspend()用于设备在切换到suspended状态时将其切换到低功耗状态。这通常包括关闭不需要的电源和时钟输入和降低设备运行频率。runtime\_resume()用来恢复runtime\_suspend()中所做的操作，将设备切换到正常使用状态。runtime\_idle()的实现是可选的，当设备驱动认为设备可以切换到suspended状态时，该函数返回0，否则返回非0值。

在实际运用中，对于嵌入式系统来说，设备状态的改变代价并不小。这会花费宝贵的CPU时间和IO等待时间。因此，Runtime PM的设计遵守一个设备只有当确信会保持低功耗一定时间的情况下才需要将其切换到suspended状态的原则，以避免将设备频繁的切换在suspended和active状态之间。为此，Runtime PM引入了autosuspend的概念。Autosuspend并非是指让设备自动进入suspended状态，而是指让设备在保持了一定的不活跃时间后才进入suspended状态。

采用autosuspend机制的设备驱动需要调用pm\_runtime\_mark\_last\_busy()来更新Runtime PM框架中该设备的不活跃时间。这通常是在设备执行完IO任务之后并且在调用pm\_runtime\_put()来减少设备usage\_count之前完成的。设备的最小不活跃时间的确定可以根据的设备采用不同的策略。设备驱动在初始化时可以通过调用pm\_runtime\_set\_autosuspend\_delay()来告诉系统这一时间。在设备运行过程中，用户空间的应用程序也可以通过/sys/devices/.../power/autosuspend\_delay\_ms这个sysfs节点来更新这一设置值。另外，设备驱动还需要调用pm\_runtime\_use\_autosuspend()来声明该设备采用autosuspend机制并且在调用Runtime PM的状态切换函数时使用\_autosuspend后缀的函数：

pm\_runtime\_suspend()要改用pm\_runtime\_autosuspend();

pm\_schedule\_suspend()要改用pm\_request\_autosuspend();

pm\_runtime\_put()要改用pm\_runtime\_put\_autosuspend();

pm\_runtime\_put\_sync()要改用pm\_runtime\_put\_sync\_autosuspend().

由于autosuspend机制下，设备完成IO任务和进入suspended状态是异步的，设备驱动就必须实现runtime\_suspend()回调函数与IO任务之间的同步。一个典型的伪代码实现如下：

foo\_read\_or\_write(struct foo\_priv \*foo, void \*data)

{

lock(&foo->private\_lock);

add\_request\_to\_io\_queue(foo, data);

if (foo->num\_pending\_requests++ == 0)

pm\_runtime\_get(&foo->dev);

if (!foo->is\_suspended)

foo\_process\_next\_request(foo);

unlock(&foo->private\_lock);

}

foo\_io\_completion(struct foo\_priv \*foo, void \*req)

{

lock(&foo->private\_lock);

if (--foo->num\_pending\_requests == 0) {

pm\_runtime\_mark\_last\_busy(&foo->dev);

pm\_runtime\_put\_autosuspend(&foo->dev);

} else {

foo\_process\_next\_request(foo);

}

unlock(&foo->private\_lock);

/\* Send req result back to the user ... \*/

}

int foo\_runtime\_suspend(struct device \*dev)

{

struct foo\_priv foo = container\_of(dev, ...);

int ret = 0;

lock(&foo->private\_lock);

if (foo->num\_pending\_requests > 0) {

ret = -EBUSY;

} else {

/\* ... suspend the device ... \*/

foo->is\_suspended = 1;

}

unlock(&foo->private\_lock);

return ret;

}

int foo\_runtime\_resume(struct device \*dev)

{

struct foo\_priv foo = container\_of(dev, ...);

lock(&foo->private\_lock);

/\* ... resume the device ... \*/

foo->is\_suspended = 0;

pm\_runtime\_mark\_last\_busy(&foo->dev);

if (foo->num\_pending\_requests > 0)

foo\_process\_requests(foo);

unlock(&foo->private\_lock);

return 0;

}

这里的重点就在于当foo\_io\_completion()完成后要求设备autosuspend的时候，设备的runtime\_suspend()回调函数(foo\_runtime\_suspend())有可能会和设备的IO任务(foo\_read\_or\_write())发生竞争现象。因此，foo\_runtime\_suspend()在让设备进入suspended状态前就必须检查是否有等待处理的IO任务。

综上所述，在实现设备的Runtime PM中，需要在设备驱动中实现runtime\_suspend(), runtime\_resume()和runtime\_idle()回调函数。在设备使用开始和结束时调用pm\_runtime\_get()和pm\_runtime\_put()来维护设备的usage\_count。在采用autosuspend机制的设备驱动中还需要决定设备的最小不活跃时间，同步runtime\_suspend()回调函数和IO任务以避免竞争。

#### Atmel MXT1664触摸屏控制器驱动的Runtime PM设计

##### **MXT1664的硬件特性**

在该嵌入式系统项目中，我们使用的是Atmel公司的MXT1664电容触摸屏控制器。MXT1664的硬件特性包括：

* maXTouch™触摸屏解决方案
  + 支持双触摸屏。每个触摸屏支持12-bit多点触摸报告和最多16个触摸并发的实时坐标跟踪
  + 支持尺寸为8到14英寸的触摸屏
* 扫描速度
  + 单点触摸最高250Hz
  + 16点触摸最高100Hz
* 响应时间
  + idle状态下的首次触摸延迟小于15ms
* 接口
  + I2C总线接口，支持400kHz快速模式以及最高1.7MHz的高速模式
  + USB 2.0标准兼容接口, full speed 12Mbps
  + Microsoft Windows 8的HID-I2C接口
* 电源
  + 最低1.8V的数字供电
  + 2.7到3.3V的模拟供电

在该系统中，我们采用MXT1664的I2C总线接口，挂载在APQ8074AB的I2C总线上。数字电源采用1.8V和3.3V，模拟电源为3.3V。通过APQ8074AB的GPIO连接芯片的控制和状态管脚如复位和中断。其接口部分原理图如下：



图 3‑3 MXT1664接口原理图

Figure 3‑3 MXT1664 interface schematics

根据芯片数据手册，MXT1664有三种工作模式: Active, Idle和Deep Sleep。其中功耗最低的工作模式是DeepSleep。各种模式下的功耗都分为三部分: 模拟电源功耗(Analog Supply, AIdd), 数字电源(Digital Supply, DIdd)和X Drive Supply(XIdd)。在Deep Sleep下，这三部分的平均耗电流分别为6.8mA, 0.08mA和0.0004mA。总的平均耗电流约为6.9mA。

##### **软件设计目标**

目前MXT1664驱动中没有实现软件上的Runtime的电源管理，当空闲时，硬件工作在MXT664的Idle模式。软件设置的空闲情况下的扫描间隔为100ms。因此，Runtime PM的软件设计目标是让MXT664在空闲时进入更低功耗的Deep Sleep模式。由于用户的触摸动作是随机的，在一次触摸之后很可能会很快发生下一个，我们采用Runtime PM的autosuspend模式来延迟一定的时间来判断硬件是否真的可以进入Runtime PM的idle状态并进而进入suspended状态。然而，MXT1664在Deep Sleep模式下时无法响应用户的触摸动作的，因此我们还需要在MXT1664进入到Deep Sleep模式之后定时地将其唤醒到Idle模式下以检测下一个用户的触摸动作。

##### **驱动流程**

基于以上目标，MXT1664驱动的Runtime PM和mxt\_runtime\_suspend()回调函数的流程可以用下图表示：

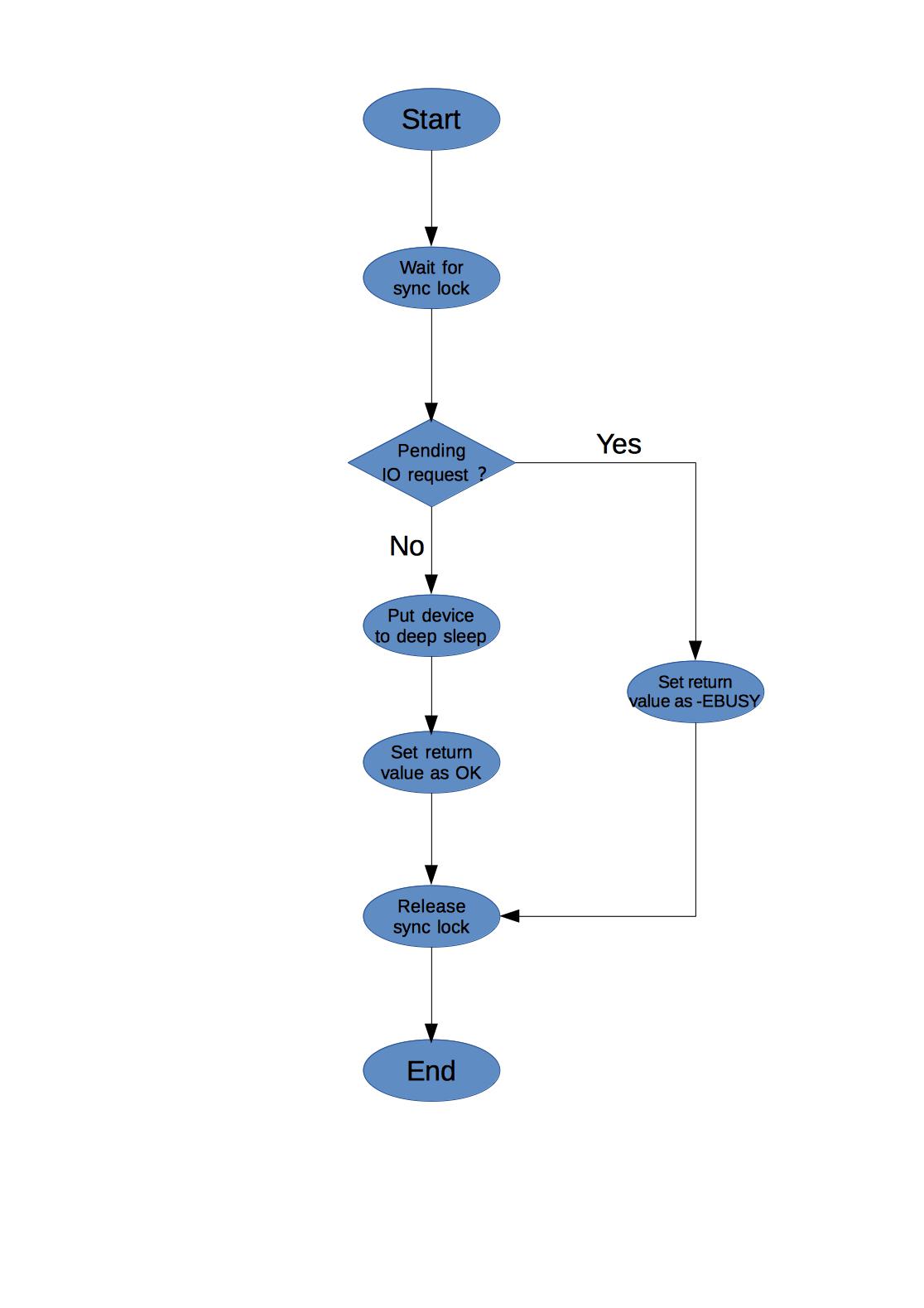
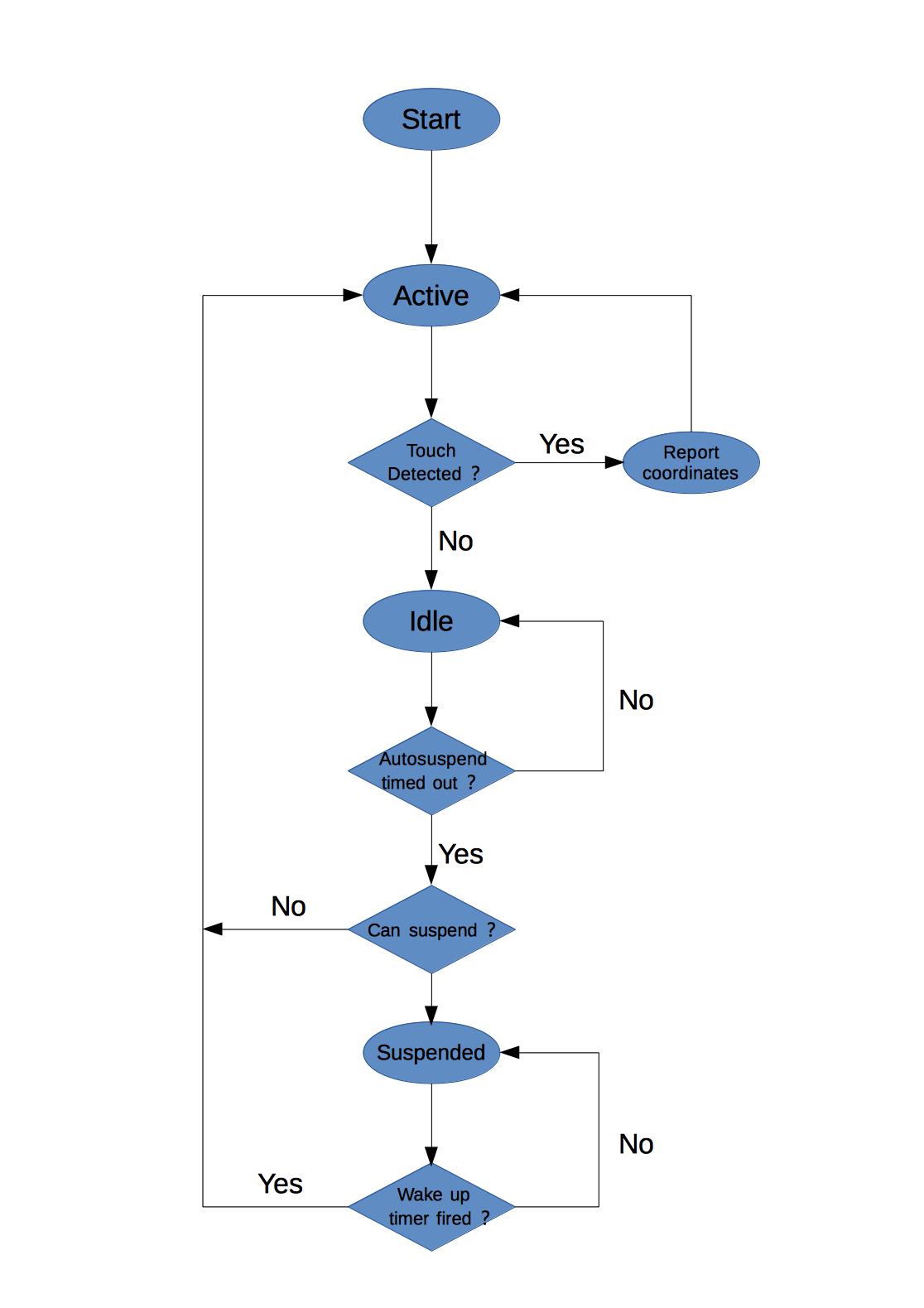


图 3‑4 MXT1664驱动流程 图 3‑5 mxt\_runtime\_suspend()流程

Figure 3‑4 MXT1664 driver flow Figure 3‑5 mxt\_runtime\_suspend() flow

mxt\_runtime\_resume()回调函数和为实现autosuspend()功能添加的mxt\_io\_complete()函数流程设计如下：

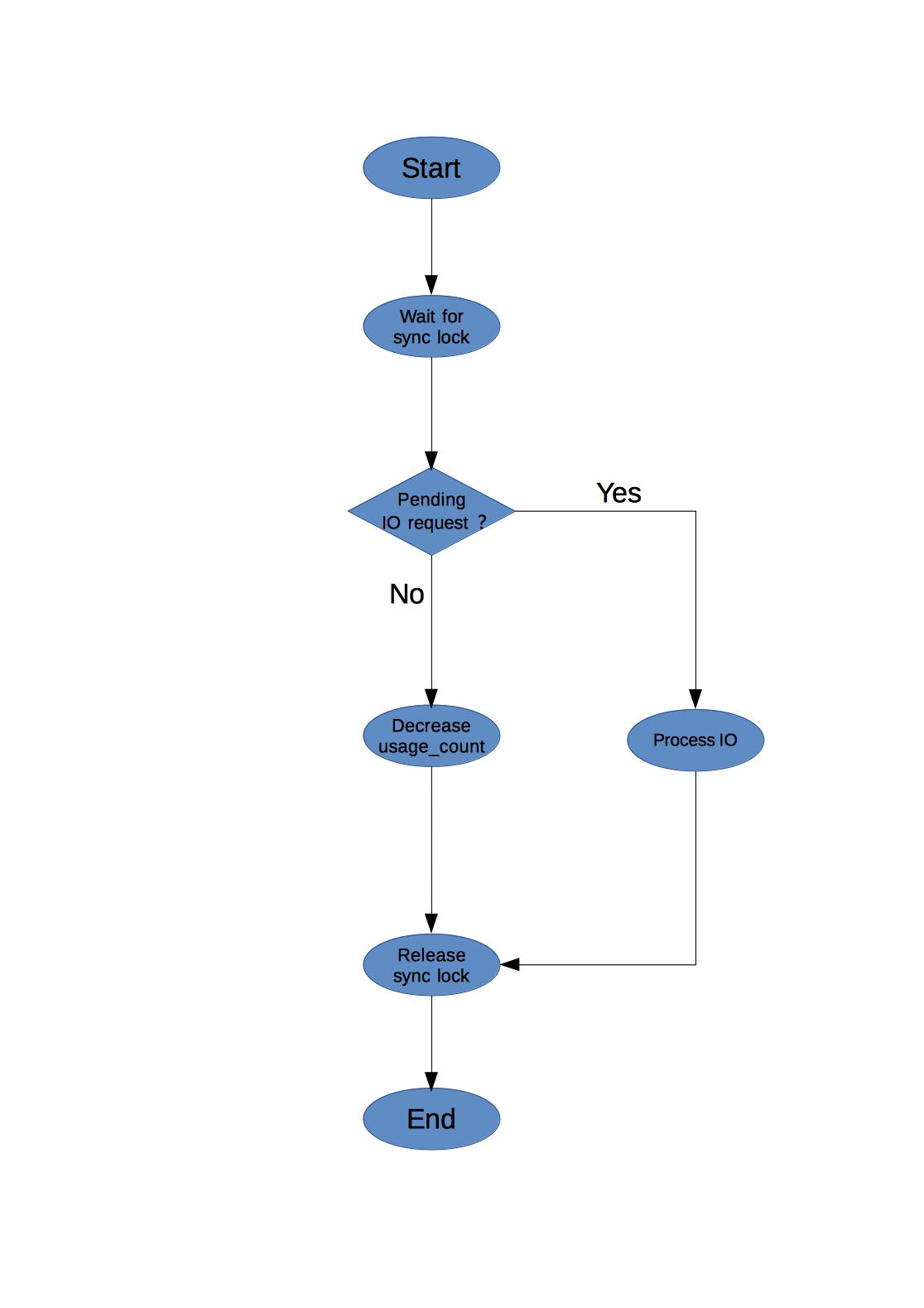
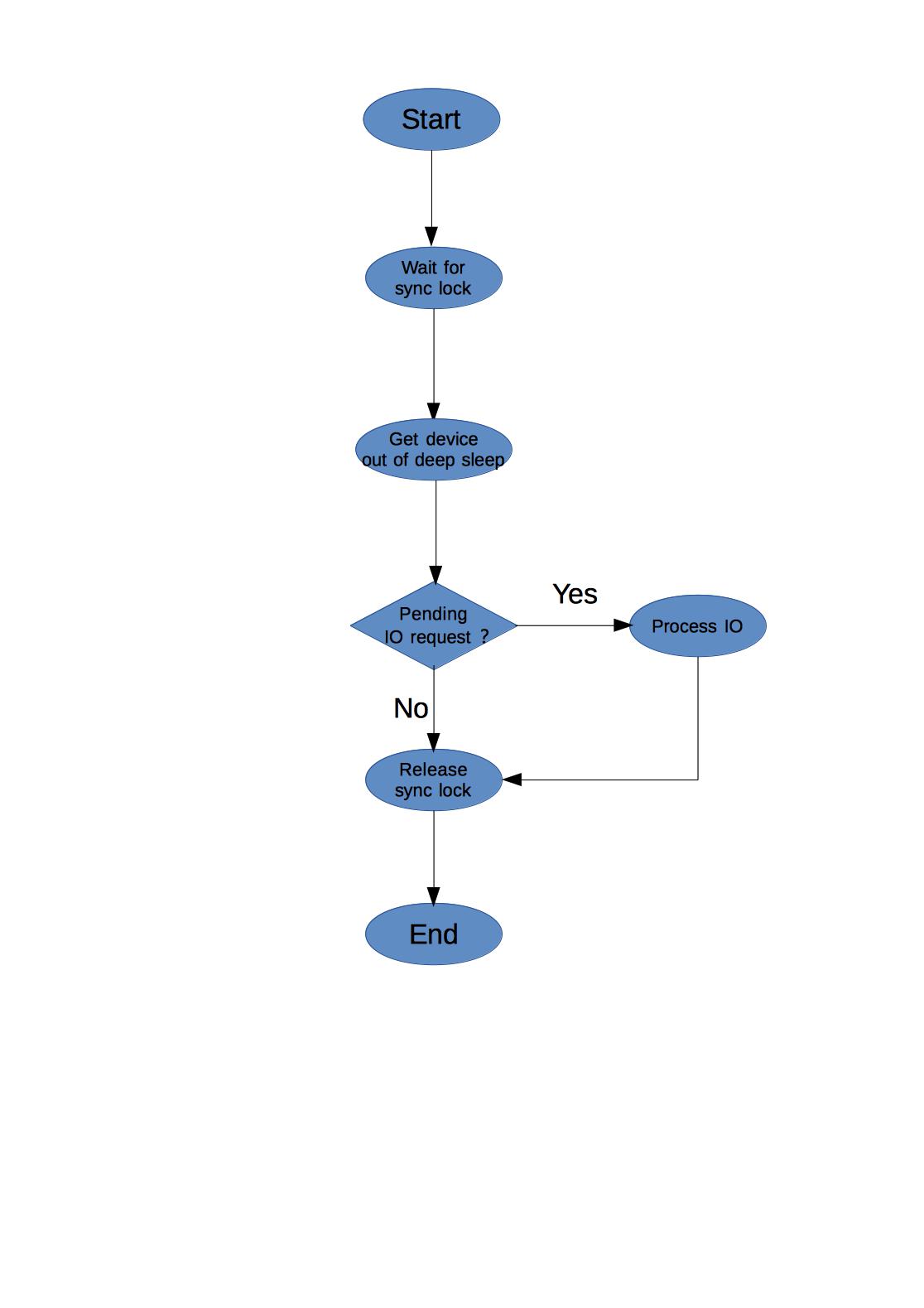


图 3‑6 mxt\_runtime\_resume()流程 图 3‑7 mxt\_io\_complete()流程

Figure 3‑6 mxt\_runtime\_resume() flow Figure 3‑7 mxt\_io\_complete() flow

## 第四章 嵌入式Linux系统电源管理的实现

## 第五章 总结与展望

## 参考文献

［1］A. Leonard Brown, Rafael J. Wysocki, Suspend-to-RAM in Linux. Proceedings of the Linux Symposium, Volume One. July, 2008.

［2］Patrick Mochel, Linux Kernel Power Management. Proceeding of the Linux Symposium. July, 2003*.*