

文章编号:1001-9081(2012)04-1176-04

doi:10.3724/SP.J.1087.2012.01176

高性能计算机系统电源设计

姚信安^{1,2*}, 宋 飞¹, 胡世平¹

(1. 国防科学技术大学 计算机学院, 长沙 410073; 2. 湖南大学 电气与信息工程学院, 长沙 410082)

(*通信作者电子邮箱 yaoxinan@sina.com)

摘要:为了满足某高性能计算机系统高效率、低成本、高可靠的供电要求,采用了12V一级母线直流分布式供电系统进行设计。介绍了计算机柜和主板的电源设计框图和工作原理,对计算主板的电压调节模块进行了详细分析,建立了基于自适应电压定位控制的电压调节模块小信号模型,分析了输出阻抗和系统控制带宽,由此得出了补偿回路设计原则,最后实测了动态响应波形。应用结果表明,电源各项技术指标完全满足该系统供电要求。

关键词:高性能计算机;分布式供电系统;电压调节模块;自适应电压定位

中图分类号: TP303.3 **文献标志码:**A

Design of power supply for high-performance computer system

YAO Xin-an^{1,2*}, SONG Fei¹, HU Shi-ping¹

(1. School of Computer Science, National University of Defense Technology, Changsha Hunan 410073, China;

2. School of Electrical and Information Engineering, Hunan University, Changsha Hunan 410082, China)

Abstract: To meet high efficiency, low cost and high reliability power requirements of high-performance computer system, 12V DC bus distributed power system was developed in this paper. The block diagram and operation principle of power supply for cabinet and motherboard were described. The voltage regulator module for processor in motherboard was analyzed in detail. Based on adaptive voltage position control, the small-signal model of voltage regulator module was presented, and output impedance and system control bandwidth were discussed. Following the proposed design guidelines of compensator, the experimental results demonstrate very good transient response. The application results show the proposed power supply can fully meet the power requirements of high-performance computer system.

Key words: high-performance computer; distributed power system; voltage regulator module; adaptive voltage position

0 引言

高性能计算机是衡量一个国家科技创新能力和综合国力的重要标志。长期以来,在全球高性能计算机TOP500排行榜中,美国一直占据榜首位置,而且在数量上也占绝对优势^[1-4]。2010年10月,国防科学技术大学研制的天河一号超级计算机系统,以峰值性能每秒4 700万亿次、持续性能每秒2 566万亿次的优异表现,双双刷新了世界超级计算机系统运算速度记录,在TOP500中首次排名世界第一^[5],使中国的高性能计算机水平向前迈进了一大步。

电源是高性能计算机系统的关键组成部分,主要功能是将交流市电转换成高质量、高效率、高可靠的直流低电压,供给处理器和内存及其他负载。随着处理器功耗越来越高,系统规模越来越庞大,千万亿次的计算机系统总功耗已经达到兆瓦级,这些都使电源设计越来越具有难度和挑战性。

近年来国内外期刊杂志和会议论文,只有少数几篇文章涉及到高性能计算机电源设计领域。文献[6]介绍了高性能计算机电源设计的关键技术,如电源架构的选择、电源可靠性的提高等,但是没有提出具体的电源设计;文献[7]则介绍了采用48 V/12 V两级母线直流分布式供电系统的某高性能计算机电源设计和稳定性分析方法,但是该系统规模有限,电源效率还有待提高。

为了提高兆瓦级高性能计算机系统的电源效率,降低研制和运营成本,本文采用了12 V一级母线直流分布式供电架构,设计和实现了某高性能计算机系统的全部电源。文中提出的电源设计和分析方法,给国内外同行提供了一定的参考价值。

1 总体设计

在某高性能计算机系统中,单板供电电压种类多达数十种,单板总功率达到800瓦,最大动态响应能力高达数百安培每微秒,单机柜最大功耗为60千瓦,系统总功耗约为6兆瓦。同时,在电压稳定性、输出电压纹波、电源效率、可靠性、维修性等方面要求非常严格。同时,要求所有插件能够热插拔和热更换。针对这些要求,采用了12 V一级母线直流分布式供电系统架构。

该系统由一百多个机柜组成,其中包含计算机柜和互通机柜。在电气上每个机柜相互独立,机柜之间采用光纤实现数据通信。图1给出了单个计算机柜的电源设计框图。

每个计算机柜分为四个独立的单元。交流380 V电压首先进入到位于机柜侧壁的电源分配单元(Power Distribution Unit, PDU)中,PDU上有数十个交流电压输出端口,通过电缆连接到一次电源模块的交流输入端口。一次电源模块的输出电压为直流12 V,最高转换效率为92%。八台一次电源模块

收稿日期:2011-09-19;修回日期:2011-11-10。 基金项目:国家863计划项目(2009AA01A128)。

作者简介:姚信安(1972-),男,江西萍乡人,副研究员,博士研究生,主要研究方向:高性能计算机电源系统;宋飞(1973-),男,江苏睢宁人,助理研究员,硕士,主要研究方向:高性能计算机电源系统;胡世平(1954-),男,湖北广济人,研究员,主要研究方向:高性能计算机电源系统设计、电磁兼容。

全部插在背板上,通过背板实现并联冗余。



图1 计算机柜电源

每块背板上插有16块计算主板、1块互连通信板、1块监控板、1块以太网络板及6个风机冷却模块。所有插件板/模块均采用双面对插方式,分布在背板两面,再通过插座从背板上取电和进行数据通信。每块背板上流过电流高达1000多安培,这对于背板的设计和制造都是巨大的挑战。

12 V电压进入到各插件板后,经过热插拔和滤波电路,再通过放置在板上的直流—直流(DC-DC)变换器转换成各种直流低电压,供给负载。

全部插件板、风机模块、一次电源模块都支持热插拔和热更换,为系统维护提供了方便。

2 计算主板电源设计

一块计算主板由4个处理器、24条内存及其他通用和专用集成电路等组成,功耗约为800瓦,需数十种直流工作电压,单路最大工作电流为100 A,最高动态响应能力为 $300 \text{ A}/\mu\text{s}$ 。每块主板由两个对称的计算节点组成,每个计算节点包括2个处理器和12条内存。图2给出了计算节点的电源框图。

两路直流12 V主电压和12 V-STBY待机电压从背板通过插座送到计算主板中,分别供给两个计算节点。经过热插拔和滤波之后,送入到各个DC-DC变换器中,转换成不同输出电压、电流和动态响应能力要求的直流低电压,供给处理器、内存条、南桥、北桥等芯片。

电源工作原理如下:12 V-STBY待机电压首先建立,产生5 V-STBY待机电压。5 V-STBY电压通过线性稳压器变成3.3 V-STBY、1.8 V-STBY、1.1 V-STBY待机电压,主板上的控制电路开始工作,进行初始化,并等待系统加电命令。当接收到系统加电命令后,控制电路按照预定的时序发出加电命令,逐步将各种DC-DC变换器加电。当各种电压都正常后,产生系统电压正常信号和复位信号,使处理器、内存条等开始正常工作,从而引导BIOS和操作系统启动。

各种DC-DC变换器设计时,根据供电电流大小和动态响应能力要求,分别采用了单相、两相和六相同步整流降压变换器。其中,处理器的供电电流为100 A,动态响应能力要求为

$300 \text{ A}/\mu\text{s}$,需按照Intel VRM11.1标准进行设计^[8]。

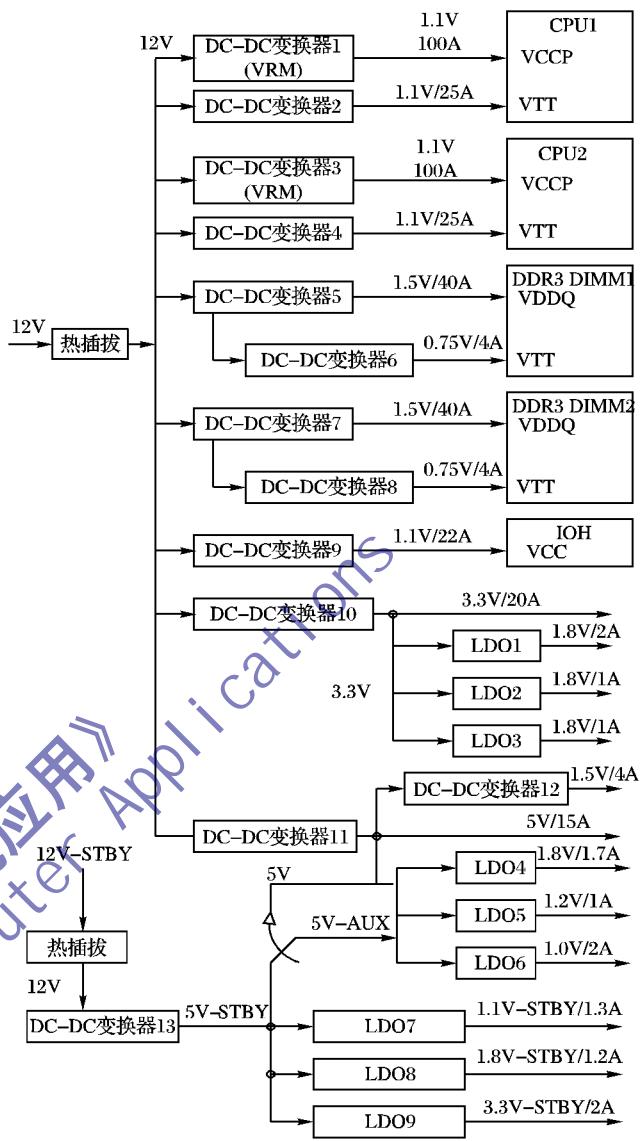


图2 计算节点电源

3 处理器电源设计

处理器的电源通常称为电压调节模块(Voltage Regulator Module, VRM)。

随着处理器技术的不断进步,供电电压已降低至0.9 V,动态响应能力已高达 $120 \text{ A}/\text{ns}$,使得VRM的设计越来越困难。传统的满足动态响应要求的方法是放置大量的电容。由于尺寸和成本的考虑,这种放置大量电容来满足动态响应要求的做法不再是一个高效的办法。因此,自适应电压定位(Adaptive Voltage Position, AVP)控制方法应运而生并得到了广泛的应用^[9-10]。其基本思想是控制VRM的输出电压,使其在轻载时输出电压比允许最高工作电压稍低一点,满载时输出电压比允许最低工作电压稍高一点。因此,在动态负载时,可以利用整个工作电压范围。图3所示即为带和不带AVP控制的VRM动态响应电压波形。显然,AVP控制将使输出滤波电容大大减少,从而减小VRM占用面积和降低成本。

根据Intel VRM11.1供电标准,采用了六相交错并联同步整流降压变换器,控制芯片采用带AVP控制的ISL6336A,电源框图如图4所示。

图 4 中, V_{in} 为输入电压, V_o 为输出电压, S1 和 S2 组成第一相, S3 和 S4 组成第二相, S11 和 S12 组成第六相, 每周期内相差 60 度导通。Lo1 ~ Lo6 为输出滤波电感, C_o 为输出滤波电容。控制芯片为 ISL6336A, MOS 管驱动芯片为 ISL6622。

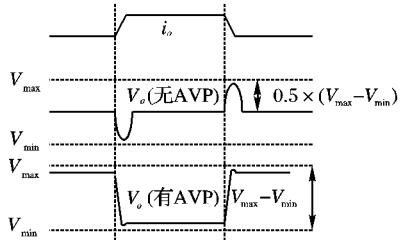


图 3 带和不带 AVP 控制的动态响应波形

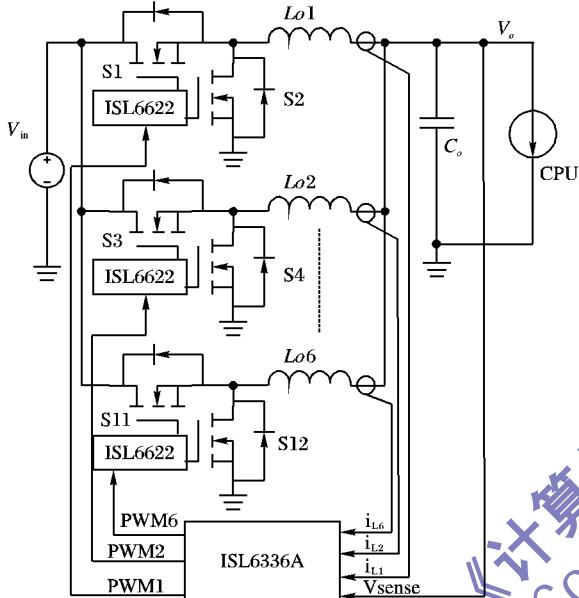


图 4 电压调节模块

多相交错并联同步整流变换器的小信号模型可简化为一个单相降压变换器。对于降压变换器来说, 实现 AVP 控制的基本方法有两种: 一种是电流控制模式, 另一种是有源下垂控制。ISL6336A 采用了有源下垂控制方法。考虑到滤波电感的等效直流电阻、滤波电容的等效串联电阻 (Equivalent Series Resistance, ESR) 等寄生参数后, 有源下垂控制单相降压变换器框图如图 5 所示。其中, 电流检测信号叠加到电压反馈信号中, 再送到补偿回路中, 这种控制方法也称作电流注入控制。

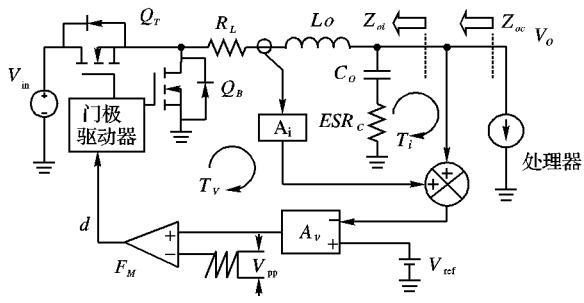


图 5 有源下垂控制单相降压变换器

根据多环设计分析方法, 有源下垂控制单相降压变换器的小信号框图如图 6 所示。其中, T_i 为电流环内环; T_v 为电压环外环; Z_{oi} 为开环输出阻抗; G_{vd} 是从控制到输出电压的传递函数; G_{id} 是从控制到电感电流的传递函数; G_u 是负载电流到电感电流的传递函数, F_M 是脉宽调制器传递函数; A_i 是电

流检测函数; A_v 是电压反馈补偿器传递函数。

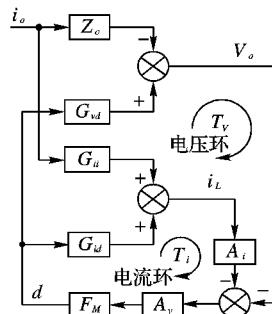


图 6 有源下垂控制单相降压变换器小信号框图

根据图 5, 可以求出:

$$Z_{oi}(s) = R_L \times \frac{(1 + \frac{s}{\omega_{ESR}}) \times (1 + \frac{s}{\omega_L})}{1 + \frac{s}{(Q \times \omega_0)} + \frac{s^2}{\omega_0^2}} \quad (1)$$

$$G_{vd}(s) = V_{in} \times \frac{1 + \frac{s}{\omega_{ESR}}}{1 + \frac{s}{(Q \times \omega_0)} + \frac{s^2}{\omega_0^2}} \quad (2)$$

$$G_{id}(s) = V_{in} \times \frac{s \times C_0}{1 + \frac{s}{(Q \times \omega_0)} + \frac{s^2}{\omega_0^2}} \quad (3)$$

$$F_M = \frac{1}{V_{pp}} \quad (4)$$

$$\omega_0 = \frac{1}{\sqrt{C_0 \times L_0}}, Q = \frac{\sqrt{C_0}}{R_L + ESR_C}, \omega_L = \frac{L_0}{R_L} \quad (5)$$

$$\omega_{ESR} = \frac{1}{ESR_C \times C_0}, \omega_L = \frac{L_0}{R_L} \quad (6)$$

其中: L_0 为输出电感, C_0 为输出电容, ESR_C 为输出滤波电容等效串联电阻, R_L 为输出滤波电感的等效直流电阻, V_{pp} 为脉宽调制器锯齿波的峰 - 峰值。

从图 6 可得出: 电流环增益为:

$$T_i = A_i \times F_M \times G_{id} \quad (7)$$

电压环增益为:

$$T_v = A_v \times F_M \times G_{vd} \quad (8)$$

比较电流模式控制和有源下垂控制的小信号框图, 可知: 有源下垂控制是电流模式控制的一个特例, 电流模式控制的设计方法完全可用于有源下垂控制电路。

如果电流环增益足够高, 则降压变换器可从二阶系统简化为一阶系统。当电流环闭环、电压环开环时, 降压变换器可看作一个理想的电流源, 如图 7 所示。此时, 输出阻抗为:

$$Z_{oi} = \frac{1}{s \times C_0} + ESR_C = \frac{1 + \frac{s}{\omega_{ESR}}}{s \times C_0} \quad (9)$$

当电压环闭环时, 闭环输出阻抗为:

$$Z_{oc} = \frac{Z_{oi}}{1 + T_2} \quad (10)$$

其中, T_2 为多环控制系统的环路增益:

$$T_2 = \frac{T_v}{1 + T_i} \quad (11)$$

输出阻抗 Z_{oc} 的转折频率为输出电容 ESR 造成的零点 ω_{ESR} 。根据恒输出阻抗设计思想, 系统环路增益 T_2 应该设计成以 -20 dB/dec 斜率穿越 0 dB , 带宽设计为 ω_{ESR} , 如图 8 所示。

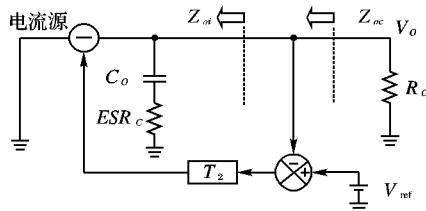


图7 电流模式控制的简化模型

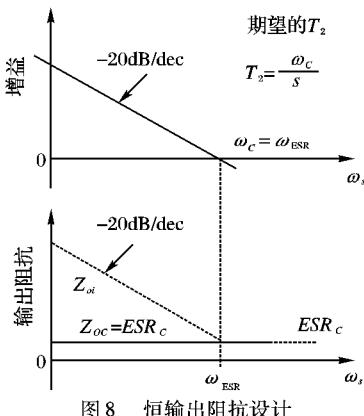


图8 恒输出阻抗设计

对于各种输出滤波电容,如有机半导体铝固体电解电容(Aluminum Solid Capacitors with Organic Semi Conductive Electrolyte, OSCON)、陶瓷电容、电解电容等,系统环路增益 T_2 期望值为:

$$T_2(s) = \frac{\omega_c}{s} \quad (12)$$

为了实现AVP控制,需按照以下几条原则和要求进行设计:

1) 电流环的带宽应该足够高,从而将降压变换器简化为一阶系统;

2) 系统环路增益设计成以-20 dB/dec斜率穿越0 dB,带宽设计为 ω_{ESR} ,相位裕度超过60°。

3) 闭环输出阻抗 Z_{oc} 小于或等于VRM的等效输出电阻 $\Delta V_o / \Delta i_o$,其中 ΔV_o 是VRM允许的最高工作电压和最低工作电压之差, Δi_o 是VRM的最大输出电流和最小输出电流之差。

如果电流环带宽足够宽,则式(11)可简化为:

$$T_2 = \frac{T_v}{1 + T_i} \approx \frac{T_v}{T_i} \quad (13)$$

根据式(2)、(3)、(7)、(8)可得出:

$$T_2(s) \approx \frac{T_v(s)}{T_i(s)} = \frac{G_{vd}(s)}{A_i \times G_{id}(s)} = \frac{1 + \frac{s}{\omega_{ESR}}}{A_i \times C_o \times s} \quad (14)$$

通常, A_i 设计成等于处理器负载线电阻 R_{droop} 。如果 T_2 带宽小于ESR零点,则式(14)可进一步简化为:

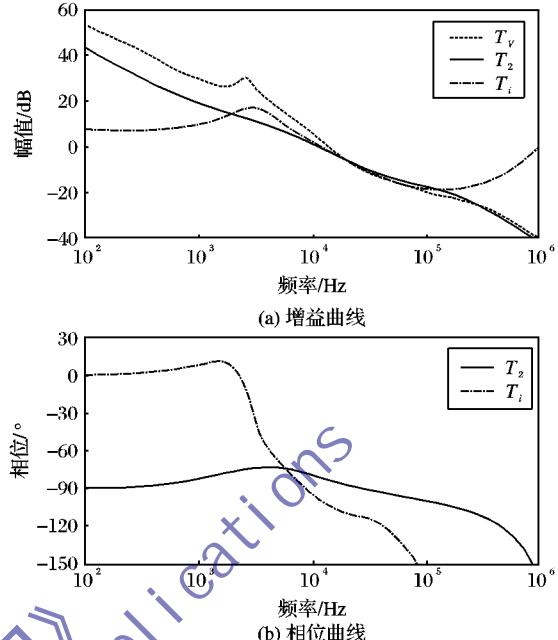
$$T_2(s) \approx \frac{1}{R_{droop} \times C_o \times s} = \frac{\omega_c}{s} \quad (15)$$

可以看出, $T_2(s)$ 正好等于期望的系统环路增益。如果电流环带宽足够高,则电压环增益 $T_v(s)$ 不影响 $T_2(s)$ 。因此,补偿电路的参数设计就变得非常简单,一般采用单极点单零点补偿电路就足够了,其传递函数如式(16) :

$$A_v = K \times \frac{1 + \frac{s}{\omega_0}}{s \times (1 + \frac{s}{\omega_p})} \quad (16)$$

其中:零点 ω_0 用来补偿系统的双极点;极点 ω_p 放在高频处,用来进一步抑制高频噪声; K 为直流增益,用来获得电流环的高带宽。计算出补偿回路参数后,即可得到 $T_v(s)$ 和 $T_i(s)$ 的传递函数。

图9示出了采用5个680 uF OSCON作为输出滤波电容时的 T_v 、 T_i 、 T_2 增益和相位曲线。可以看出,采用OSCON电容时, T_v 、 T_i 、 T_2 三个环路的带宽几乎相同,均为电容ESR引起的零点 ω_{ESR} 。

图9 OSCON输出电容时的 T_v 、 T_i 、 T_2 增益和相位曲线

根据以上理论分析,采用ISL6336A设计了一个12 V输入、1.1V/100 A输出的VRM。图10给出了实测动态响应电压电流波形。可以看出,动态响应波形符合AVP控制思想,因此可以满足处理器的供电要求。

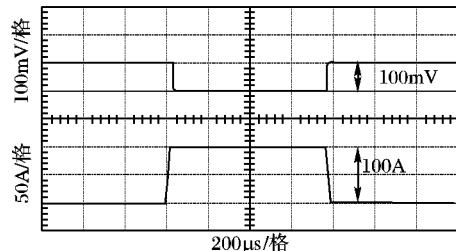


图10 实测VRM动态响应波形

4 结语

本文采用12 V一级母线直流分布式供电系统,设计和实现了某高性能计算机系统的电源。介绍了计算机柜和主板的电源设计框图和工作原理,并重点介绍了处理器电压调节模块的设计方法。基于自适应电压定位控制思想,设计了采用ISL6336A的六相交错并联同步整流电压调节模块。建立了基于有源下垂控制的电压调节模块小信号模型,求出了输出阻抗、环路增益及其他传递函数,由此得出了电压反馈补偿电路的设计原则,最后实测了电压调节模块的动态响应电压电流波形。测量结果表明,电压调节模块设计很好地满足了处理器供电要求。电源安装到主板和机柜以后,实测了各项电源技术指标,完全满足系统供电要求,能耗比参数高达600 MFlops/W,实现了高效节能、低成本、高可靠的电源设计目标。该高性能计算机系统目前已经装机运行,峰值速度和实测性能达到了国际领先水平。

参考文献:

- [1] Supercomputer TOP500 organization. Top 500 list (November 2006) [EB/OL]. [2011-09-20]. http://www.top500.org/static/lists/2006/11/top500_200611.xls. (下转第1187页)

原因归结为是由于“横电场”数据的不完整,因为菲涅耳研究所的科研人员只测量了垂直于径向的那一部分数据^[5]。如表2所示,对于单个频率的“横电场”数据,由于采用了更复杂的矢量积分算子,对比源反演算法和乘法正则化的对比源反演算法的计算时间约为8 h。

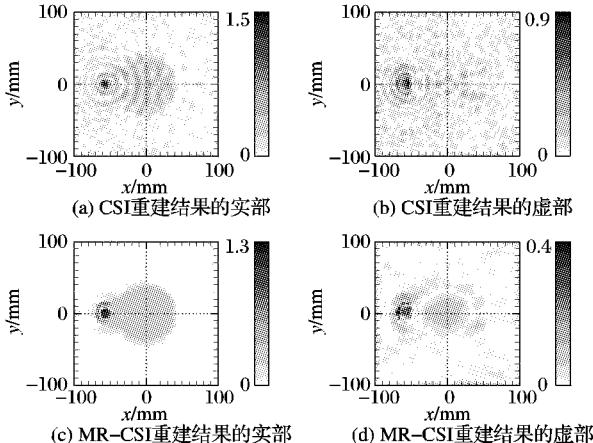


图5 CSI 和 MR-CSI 在频率 10 GHz 下对“横电场”实测数据的重建结果

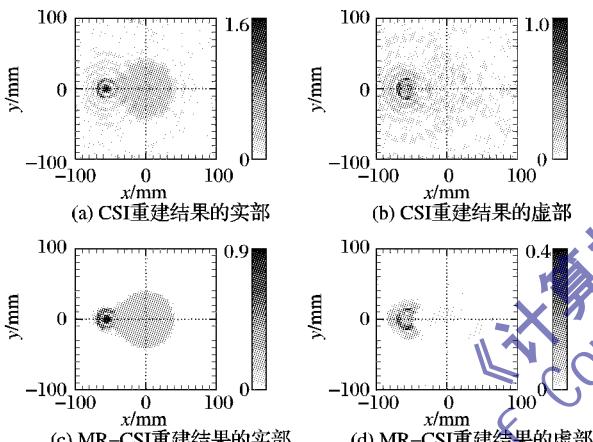


图6 采用CF方法的CSI和MR-CSI对“横电场”实测数据的重建结果

4 结语

从重建结果可以看到,对比源反演算法取得的重建结果明显优于线性多重信号分类算法的重建结果。与对比源反演算法相比,多重信号分类算法的最大优势是能够实现“准实时成像”。但是该线性算法的劣势就是重建结果的图像质量要相对差很多,而且图像所包含的有关目标的信息也有限。本文主要提出了采用时间反演的多重信号分类算法、在对比源反演算法基础上的扩展算法(如乘法正则化和并行频率方

法)对多频、多收发的实测微波数据成像的具体方法,通过比较这些线性和非线性算法的重建结果,表明扩展后的对比源反演算法是精确、有效的成像算法。

参考文献:

- [1] MARKLEIN R, MAYER K, HANNEMANN R, et al. Linear and nonlinear inversion algorithms applied in nondestructive evaluation [J]. Inverse Problems, 2002, 18(6): 1733–1759.
- [2] ABUBARKAR A, van den BERG P M. Iterative forward and inverse algorithms based on domain integral equations for three dimensional electric and magnetic objects [J]. Journal of Computational Physics, 2004, 195(1): 236–262.
- [3] YU C, SONG L P, LIU Q H. Inversion of multi-frequency experimental data for imaging complex objects by a DTA-CSI method [J]. Inverse Problems, 2005, 21(6): S165–S178.
- [4] GEFFRIN J M, SABOUREUX P, EYRAUD C. Free space experimental scattering database continuation: Experimental set-up and measurement precision [J]. Inverse Problems, 2005, 21(6): S117–S130.
- [5] MIAO J. Linear and nonlinear inverse scattering algorithms applied in 2-D electromagnetics and elastodynamics [M]. Kassel: Kassel University Press, 2008: 1–175.
- [6] CHENEY M. The linear sampling method and the MUSIC algorithm [J]. Inverse Problems, 2001, 17(4): 591–595.
- [7] PARK W, LESSELIER D. MUSIC-type imaging of a thin penetrable inclusion from its multi-response matrix [J]. Inverse Problems, 2009, 25(7): 1–34.
- [8] MARENKO E A, GRUBER F K, SIMONETTI F. Time-reversal MUSIC imaging of extended targets [J]. IEEE Transactions on Image Processing, 2007, 16(8): 1967–1984.
- [9] KLEINMAN R E, van den BERG P M. A contrast source inversion method [J]. Inverse Problems, 1997, 13(6): 1607–1620.
- [10] EGGER H, LEITÄO A. Nonlinear regularization methods for ill-posed problems with piecewise constant or strongly varying solutions [J]. Inverse Problems, 2009, 25(11): 1–19.
- [11] BACHMAYR M, BURGER M. Iterative total variation schemes for nonlinear inverse problems [J]. Inverse Problems, 2009, 25(10): 1–26.
- [12] ABUBARKAR A, HU W, van den BERG P M, et al. A finite-difference contrast source inversion method [J]. Inverse Problems, 2008, 24(6): 1–17.
- [13] MIAO J, MARKLEIN R, LI J. Application of the linear and nonlinear inversion algorithms on two-dimensional experimental electromagnetic data [C]// the International Conference on Microwave Technology and Computational Electromagnetics. Beijing: [s. n.], 2009: 296–299.

(上接第 1179 页)

- [2] Supercomputer TOP500 organization. Top 500 list (November 2007) [EB/OL]. [2011-09-20]. http://www.top500.org/static/lists/2007/11/top500_200711.xls.
- [3] Supercomputer TOP500 organization. Top 500 list (November 2008) [EB/OL]. [2011-09-20]. http://www.top500.org/static/lists/2008/11/top500_200811.xls.
- [4] Supercomputer TOP500 organization. Top 500 list (November 2009) [EB/OL]. [2011-09-20]. http://www.top500.org/static/lists/2009/11/top500_200911.xls.
- [5] Supercomputer TOP500 organization. Top 500 list (November 2010) [EB/OL]. [2011-09-20]. http://www.top500.org/static/lists/2010/11/top500_201011.xls.
- [6] 胡世平, 姚信安, 宋飞. 高性能计算机电源系统设计的关键技

- 术[J]. 计算机工程与科学, 2006, 28(5): 136–140.
- [7] 姚信安, 胡世平, 宋飞, 等. 高性能计算机电源系统的设计与实现[J]. 电力电子技术, 2008, 42(2): 40–42.
- [8] Intel. Voltage regulator module (VRM) and enterprise voltage regulator-down (EVRD) 11.1 design guidelines [EB/OL]. [2011-09-20]. <http://www.intel.com/Assets/PDF/designguide/397898.pdf>.
- [9] ZHANG M T. Powering Intel pentium 4 generation processors [C]// Electrical Performance of Electronic Packaging. Hillsboro: IEEE, 2001: 215–218.
- [10] 袁伟, 张军明, 钱照明. 一种混合式自适应电压定位控制策略及 12 V 电压调节模块拓扑[J]. 电工技术学报, 2010, 25(10): 115–121.