

利用企查查和百度地图挖掘企业信息实验数据集—— 以珠三角地区为例（1980–2020）

王培涵¹², 查瑞波^{2*}, 黄金川^{134*}, 黄悦², 杜书滢², 许进谿²

1. 中国科学院地理科学与资源研究所, 北京 100101; 2. 福建师范大学文化旅游与公共管理学院, 福州 350108; 3. 中国科学院区域可持续发展分析与模拟重点实验室, 北京 100101; 4. 中国科学院大学资源与环境学院, 北京 100049

摘要: 利用“企查查”平台数据和“百度地图”平台数据, 通过数据挖掘技术可以得到有关企业的时空分布数据和信息。作为电子信息产业的基础和改造传统产业的核心, 集成电路产业的发展是推动战略性新兴产业不断发展的关键。珠三角地区是我国集成电路产业的集聚区之一, 同时又是我国大陆改革开放的前沿阵地, 对于该地区集成电路企业的时空分布研究, 有利于加深对我国集成电路产业发展现状的了解, 以及对其他地区集成电路产业发展提供前车之鉴。本文基于企查查企业信息发布平台以及百度地图开放平台进行数据的采集、处理和汇总, 得到珠三角地区集成电路企业地理位置、成立时间、经营状态等数据, 进一步采用核密度分析方法探究珠三角集成电路企业分布时空变化格局。数据集以.shp 格式存储, 由 7 个数据文件组成, 数据量为 45.8 MB (压缩为 1 个文件, 710 KB)。

关键词: 集成电路; 珠三角地区; 企业; 时空分布

DOI: <https://doi.org/10.3974/geodp.2022.04.14>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2022.04.14>

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志(中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2022.10.07.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2022.10.07.V1>。

1 前言

集成电路产业是国民经济的先导性产业, 是个人电脑、互联网、大数据、智能手机等新一代信息技术产业发展的核心前提, 融入了人民生活各个方面。在当前国际环境下集成电路产业重要的战略性作用更为突出, 因其对全球化分工的高度依赖, 集成电路产业已经成为某些国家影响国际政治和经济的手段。一个国家或地区若能够在集成电路产业链

收稿日期: 2022-11-01; 修订日期: 2022-12-15; 出版日期: 2022-12-24

基金项目: 国家自然科学基金(42201267); 教育部(18YJJCZH007)

*通讯作者: 查瑞波, 福建师范大学文化旅游与公共管理学院, rbzha@fjnu.edu.cn

黄金川, 中国科学院地理科学与资源研究所, huangjc@igsrr.ac.cn

数据引用方式: [1] 王培涵, 查瑞波, 黄金川等. 利用企查查和百度地图挖掘企业信息实验数据集——以珠三角地区为例(1980–2020)[J]. 全球变化数据学报, 2022, 6(4): 631–637. <https://doi.org/10.3974/geodp.2022.04.14>. <https://cstr.escience.org.cn/CSTR:20146.14.2022.04.14>.

[2] 王培涵, 查瑞波, 黄悦等. 利用企查查和百度地图挖掘企业信息实验数据集——以珠三角地区为例(1980–2020)[J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.10.07.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2022.10.07.V1>.

条中占据重要位置,则不仅可以在世界市场中获得巨大收益,同时在国际竞争中拥有更大话语权。但如果过于依赖国外集成电路产业则可能在关键领域受制于人,基于此,我国集成电路产业的发展日益迫切^[1,2]。

中国集成电路产业发展开始于20世纪70年代后期,经历了近30年的摸索在2009年进入了高速发展阶段,2009–2019年中国集成电路产业平均每年销售额增长645.3亿元,同时期中国也成为世界最大的半导体市场。虽然发展迅速、市场广阔,但是中国集成电路产业整体呈现实力较弱、产品较初级、高端产品依赖进口、贸易逆差大,自主可控的产业链尚未形成的特点,严重地威胁着中国集成电路产业及相关产业的安全,制约着产业的发展壮大^[3]。近年来国家对于集成电路产业也愈发重视,2021年《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》中明确指出,集成电路是事关国家安全和全局的基础核心领域,需制定实施战略性科学计划和科学工程,推动集成电路产业创新发展。当前,有关集成电路产业的研究多集中于产业链创新^[4,5]、发展脉络与格局^[3,6]、产业协同^[7]等领域,对于集成电路的空间分布研究较少^[8]。

珠三角地区是我国集成电路产业的集聚区之一^[9],同时又是我国大陆改革开放的前沿阵地^[3],对于该地区集成电路企业的时空分布研究,有利于加深对我国集成电路产业发展现状的了解,以及对其他地区集成电路产业发展提供前车之鉴^[10]。为此,本数据集提供了珠三角地区集成电路企业空间分布数据,该数据包括集成电路企业的成立时间、所属县市区、产业分工等信息,可为相关研究提供基础数据支撑。

2 数据集元数据简介

《利用企查查和百度地图挖掘企业信息实验数据集——以珠三角地区为例(1980–2020)》^[11]的名称、作者、地理区域、数据年代、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表1。

3 数据研发方法

本研究的基础数据来源于企查查网站¹,以“集成电路”和“芯片”为关键词从该网站上获取了珠三角地区9个地市的集成电路企业名录共14,195个企业,记录了每个企业的名称、注册地、成立时间等信息。由于企查查网站没有企业地理坐标,故根据企业地理位置利用百度地图开放平台²获取相应地理坐标。通过人工进行企业网站和电话问询确定企业在产业链的分工。

4 数据结果

4.1 数据集组成

本数据集包含了企查查网站上珠三角九个地市的集成电路企业名录,共收集了14,195个企业的信息。数据集以空间点数据进行存储,存储为.shp格式数据文件。记录的属性字

¹ 企查查网. <https://www.qcc.com/>.

² 百度地图开放平台. <https://lbsyun.baidu.com/>.

段如表2所示。

表1 《利用企查查和百度地图挖掘企业信息实验数据集——以珠三角地区为例（1980-2020）》元数据简表

条 目	描 述
数据集名称	利用企查查和百度地图挖掘企业信息实验数据集——以珠三角地区为例（1980-2020）
数据集短名	PRD_IC_1980-2020
作者信息	王培涵，中国科学院地理科学与资源研究所，福建师范大学文化旅游与公共管理学院，wph1996@126.com 查瑞波，福建师范大学文化旅游与公共管理学院，rbzha@fjnu.edu.cn 黄悦，福建师范大学文化旅游与公共管理学院，hyueyuey@163.com 杜书滢，福建师范大学文化旅游与公共管理学院，huoying_ll@126.com 许进鎔，福建师范大学文化旅游与公共管理学院，xjr9018@163.com
地理区域	珠江三角洲
数据年代	1980-2020
数据格式	.shp
数据量	45.8 MB（710 KB 压缩后）
数据集组成	由 14195 条企业数据构成
基金项目	国家自然科学基金项目（42201267）；教育部（18YJ CZH007）
出版与共享服务	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、通过《全球变化数据仓储电子杂志（中英文）》发表的实体数据集和通过《全球变化数据学报（中英文）》发表的数据论文。其共享政策如下：（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则，即从本数据集中摘 ^[12]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

表2 数据集中的属性字段表

序号	字段名	字段内容举例
1	名称_中文	深圳市航顺集成电路技术研发有限公司、广东利扬集成电路测试股份有限公司、深圳衡宇集成电路科技有限公司……
2	经度	113.278,241,4、114.277,798,2、113.925,025,9……
3	纬度	22.839,827,95、22.722,734,04、22.495,764……
4	经营状态_中文	在业、存续、注销……
5	成立时间	2013/8/15、2017/5/12、2013/10/14……
6	城市名_中文	深圳、广州、惠州……
7	县区名_中文	宝安区、黄浦区、香洲区……
8	企业性质_中文	有限责任公司、股份有限公司（上市）、其他
9	产业分工_中文	设计、制造、封测……

4.2 数据结果

4.2.1 空间分布

从县区市层面看，如图 1 所示，集成电路企业较多的地区为深圳市各区、东莞市、惠州市惠城区、广州市黄浦区、佛山市顺德区、中山市以及珠海市香洲区。其余地区企业数量较少。总体来看呈现以深圳为核心，沿珠江入海口呈倒 U 型集中分布的特征。

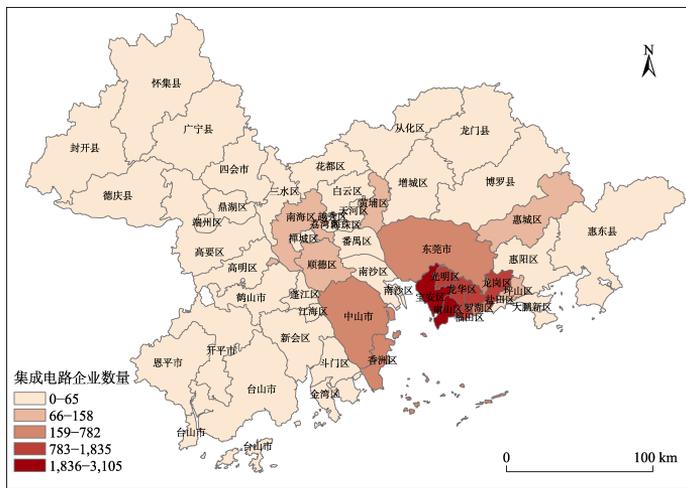


图 1 珠三角地区集成电路企业数量空间分布图
(依据审图号 GS(2019)1822 号标准地图制作)

利用 ArcGIS 软件计算珠三角地区集成电路企业空间自相关程度，以珠三角各县市区为基本单元，计算个单元集成电路企业数量的空间自相关关系，得到 Moran's I 指数为 0.793，z 值为 10.100，p 值为 0，说明珠三角集成电路企业在县区尺度存在较高的空间集聚。为进一步分析集成电路企业在县区内部的分布情况，利用 ArcGIS 软件计算珠三角地区集成电路企业核密度（图 2），从图中可以看出各县区内部分布也呈明显的集聚趋势，深圳市核

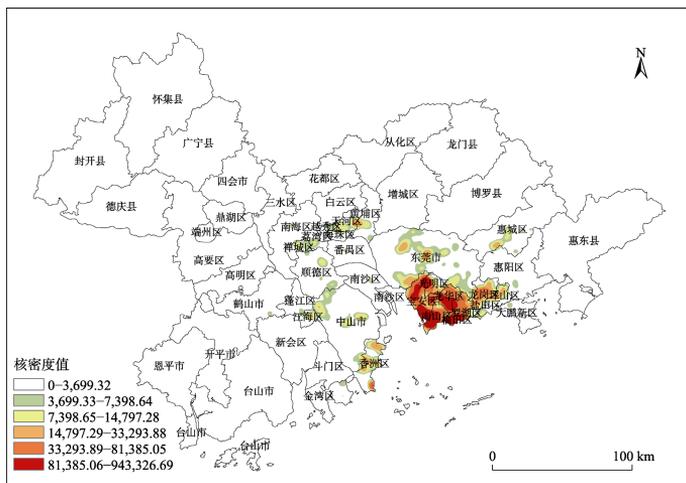


图 2 珠三角地区集成电路企业分布核密度图
(依据审图号 GS(2019)1822 号标准地图制作)

密度明显高于其他区域，另外在东莞市中部以及与深圳市接壤的地区、惠城区中南部、广州市的天河区和黄埔区交界处、禅城区中东部、珠海市香洲区东南部以及与中山市接壤地区等都有一定程度上的集中，核密度最高值位于宝安区北部工业园区。

4.2.2 时间演变

为探究珠三角地区集成电路企业成立发展过程，根据企业成立时间将企业分为8个时间段：1980-1984年、1985-1999年、1990-1994年、1995-1999年、2000-2004年、2005-2009年、2010-2014年、2015-2020年进行作图，如图3所示。

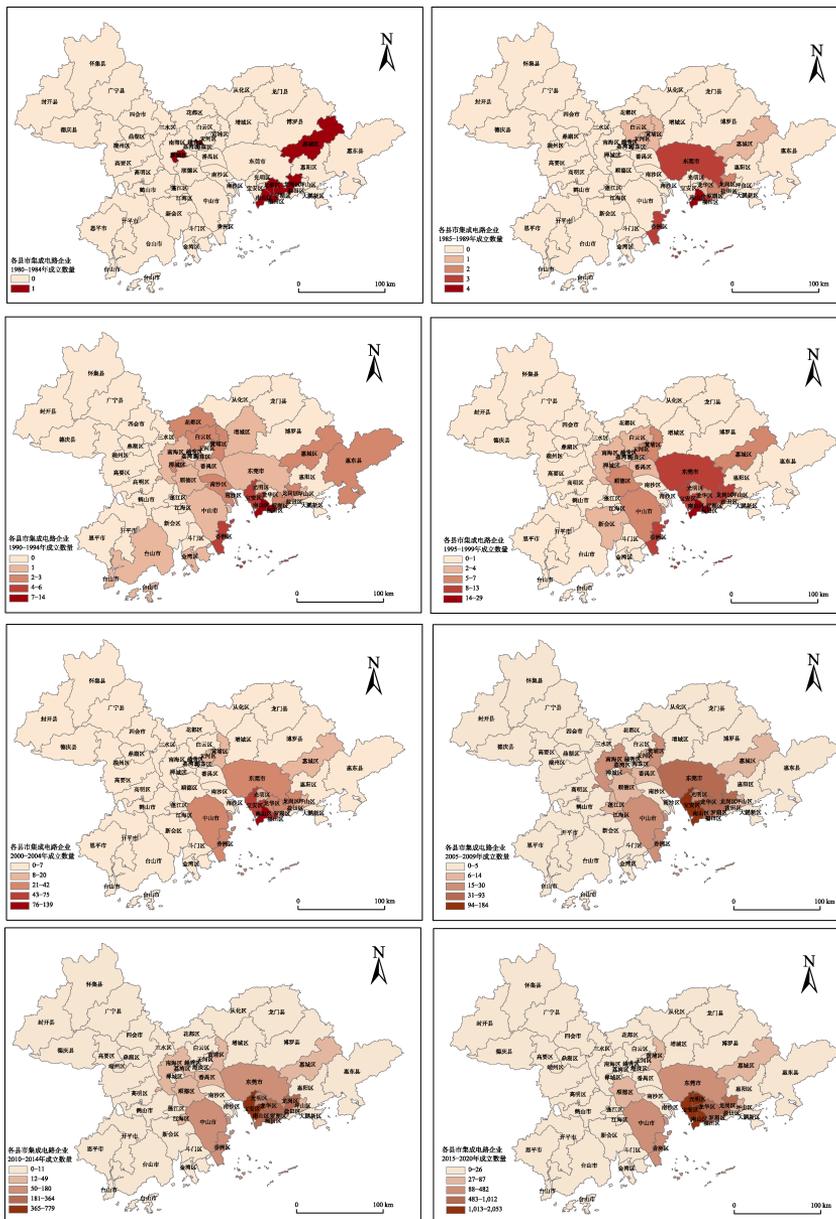


图3 1980-2020年每5（/6）年珠三角各县市集成电路企业成立数量图
（依据审图号GS(2019)1822号标准地图制作）

1980–1984 年间, 集成电路企业零星出现。深圳市、广州市、佛山市以及惠州市开始出现集成电路企业, 总体数量较少, 从县区分布来看各个县区数量差异不大。

1985–1989 年间, 新成立企业主要集中在珠江口东侧。珠海市香洲区、广州市黄埔和从化市也有集成电路企业成立, 总体数量有所增长, 深圳市南山区是增加数量最多的区。

1990–1994 年间, 企业成立数量增多, 范围也逐渐扩大。江门市和中山市也有集成电路企业成立, 深圳市企业数量逐渐与其他地市拉开差距。

1995–1999 年间, 企业逐渐向珠江口沿岸集中。深圳市和东莞市在成立数量上逐渐与其他城市拉开差距。

2000–2004 年间, 深圳市增速继续加快, 南山区企业增加数量突破 100 个, 宝安区、福田区和龙岗区紧随其后形成第二梯队。

2005–2020 年间, 集成电路企业增加迅速, 沿珠江口呈倒 U 型分布格局逐渐形成, 深圳增加最多。

总体来看, 二十世纪 80 年代开始, 珠三角地区个别城市率先成立集成电路企业, 最初集中在珠江口东岸; 二十世纪 90 年代企业成立的数量和覆盖城市逐渐增加; 进入二十一世纪, 珠江口沿岸地区集成电路企业数量增加迅速并逐渐形成沿珠江口的倒 U 型企业分布格局, 其中深圳企业成立数量远超其他城市。

4.2.3 企业产业分工

随着集成电路产业的发展, 出现了成熟的产业分工, 包括上游的支持企业, 核心的设计、制造、封测企业以及下游的需求企业, 同时有一些企业参与多个分工。为探究珠三角地区集成电路企业分工状况, 根据企业自述以及网站和电话问询, 划分企业在产业链条的位置, 将企业划分为支持企业、设计、制造、封测、需求企业、设计与制造、制造与封测、设计制造封测 8 种产业分工, 如表 3 所示。

表 3 集成电路企业产业分工统计表

企业分工	支持企业	核心企业						需求企业
		设计	制造	封测	设计与制造	制造与封测	设计制造封测	
企业数量/个	23	257	41	68	22	3	12	13766
占比/%	0.16%	1.81%	0.29%	0.48%	0.15%	0.02%	0.08%	96.98%

珠三角地区集成电路企业约 97% 以上为集成电路产业下游需求企业共计 13,766 家, 多为利用成品集成电路进行产品的制造, 以家电制造、LED 灯饰、智能卡等为主。上游支持企业十分稀少, 仅占 0.16%。核心企业共计 403 家, 约占 2.84%, 其中集成电路设计企业最多占 1.81%, 集成电路封测企业次之占 0.48%, 集成电路制造企业占 0.29%。

5 结论

面对愈发复杂的国际形势珠三角地区作为我国集成电路产业重点集聚地, 对其集成电路企业的时空演变规律的研究具有重要意义。本数据集收集了 1980–2020 年间珠三角地区成立的集成电路企业名录, 共计 14,195 个企业, 包含了每个企业的名称、经营状态、所在

地、产业分工、成立时间、注册资本等详细属性信息。通过研究发现，在空间上，集成电路企业空间分布呈现以深圳为核心，沿珠江入海口呈倒U型集中分布的特征；在时间上，二十世纪80年代开始个别城市出现集成电路企业，而后各个城市均有集成电路企业成立，进入二十一世纪特别是2010年后集成电路企业成立数量迅速增加，其中深圳企业成立数量远超其他城市；在产业分工上，数量最多的是下游需求企业，数量最少的是上游支持企业，在集成电路核心企业中设计企业数量最多，制造类企业较少。本数据集所提供的企业信息尚有在较大的数据信息挖掘潜力，望为珠三角地区集成电路产业研究提供参考和数据基础。

作者分工：王培涵对数据集的开发和研究做了总体设计，并撰写了数据论文；黄悦、杜书滢和许进铨收集和整理了本数据集；查瑞波和黄金川对数据集和论文进行了修改。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] 朱贻玮. 集成电路产业50年回眸[M]. 北京: 电子工业出版社, 2016.
- [2] 刘雯, 马晓辉, 刘武. 中国大陆集成电路产业发展态势与建议[J]. 中国软科学, 2015(11): 186–92.
- [3] 王阳元. 掌握规律, 创新驱动, 扎实推进中国集成电路产业发展[J]. 科技导报, 2021, 39(3): 31–51.
- [4] 曲永义, 李先军. 创新链赶超: 中国集成电路产业的创新与发展[J]. 经济管理, 2022, 44(9): 5–26.
- [5] 吴菲菲, 韩朝曦, 黄鲁成. 集成电路产业研发合作网络特征分析——基于产业链视角[J]. 科技进步与对策, 2020, 37(8): 77–85.
- [6] 王耀南, 廖蕾. “中国芯”破局之路如何走[J]. 人民论坛, 2022(17): 60–63.
- [7] 孙琴, 刘戒骄, 胡贝贝. 中国集成电路产业链与创新链融合发展研究[J]. 科学学研究, 2022(10): 1–20.
- [8] 任亚文, 杨宇. 珠三角地区半导体产业布局特征及其区位关联模式[J]. 地理科学进展, 2022, 41(9): 1622–1634.
- [9] 叶玉瑶, 吴康敏, 张虹鸥等. 珠三角新晋跨境制造业企业地理集聚与区位选择[J]. 地理科学进展, 2019, 38(10): 1583–1595.
- [10] 贺灿飞, 王文宇, 郭琪. 中国工业地理学的传承发展与转型[J]. 地理学报, 2021, 76(8): 1815–1834.
- [11] 王培涵, 查瑞波, 黄悦等. 利用企查查和百度地图挖掘企业信息实验数据集——以珠三角地区为例（1980–2020）[J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.10.07.V1>. <https://cstr.science.org.cn/CSTR:20146.11.2022.10.07.V1>.
- [12] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017年更新).