

扩展现实设备芯片需求白皮书

White Paper on extended reality Device Chip Requirements

(2023 年)

中国通信标准化协会

扩展现实（XR）产业及标准推进委员会

2023 年 5 月

版权声明

本白皮书版权属于扩展现实（XR）产业及标准推进委员会和中国通信标准化协会，并受法律保护。转载、摘编或利用其他方式使用本白皮书文字或者观点的，应注明“来源：扩展现实（XR）产业及标准推进委员会和中国通信标准化协会”，违者追究法律责任。

指导单位

中国通信标准化协会 扩展现实（XR）产业及标准推进委员会

参编单位

安谋科技（中国）有限公司	杭州灵伴科技有限公司(Rokid)
紫光展锐(上海)科技有限公司	OPPO 广东移动通信有限公司
创通联达智能技术有限公司	瑞芯微电子股份有限公司
维沃移动通信有限公司	中国移动通信集团终端有限公司
博鼎实华（北京）技术有限公司	上海市多媒体行业协会

感谢以下机构对本白皮书的编写提出宝贵建议

翱捷科技股份有限公司	歌尔股份有限公司
优奈柯恩(北京)科技有限公司	雷鸟创新技术（深圳）有限公司
深圳创维新世界科技有限公司	晨讯科技(上海)有限公司

玩出梦想（上海）信息科技有限公司等

编写组

毛卫洋 伍赛 陈 锴 陈翊翔 任 远 刘 畅 李丛蓉 钟宝星

马 超 王 真 康 峰 杨冬笋 来航曼 吴传龙 符谋政 林怀亮

董丙银 刘 峰 倪 茂 彭 程 郑海霞 端木海婴 李 技 周晓萌

朱勇旭 黄德胜 彭华成

扩展现实产业及标准推进委员会
(TC625)

目录

前 言	I
第一章 扩展现实产业概述	1
第二章 现有热门扩展现实设备方案概览	8
2.1 VR 一体机概览.....	8
2.1.1 VR 一体机热门设备参数.....	8
2.1.2 VR 一体机国内设备介绍.....	9
2.2 主机式 VR 概览	11
2.2.1 主机式 VR 热门设备参数	12
2.2.2 主机式 VR 国内设备介绍	12
2.3 AR 一体机概览.....	13
2.3.1 AR 一体机热门设备参数.....	14
2.3.2 AR 一体机国内设备介绍.....	15
2.4 AR 分体机概览.....	16
2.4.1 AR 分体机热门设备参数.....	17
2.4.2 AR 分体机国内产品介绍.....	18
第三章 扩展现实设备底层技术需求	21
3.1 XR 基本组件.....	21
3.2 主要参数指标要求	24
3.2.1 延迟	25
3.2.2 视场角	26
3.2.3 角分辨率	27
3.2.4 续航及重量	29
3.3 核心算法技术.....	30
3.3.1 DoF 技术.....	30
3.3.2 ATW/ASW 技术.....	31
3.3.3 眼动追踪	33

3.3.4 注视点渲染	34
3.3.5 透视技术	35
3.3.6 手势追踪技术	37
3.3.7 手柄追踪技术	39
第四章 扩展现实设备底层芯片需求	41
4.1 核心计算芯片能力	41
4.1.1 计算能力	43
4.1.2 无线通信能力	48
4.1.3 芯片接口	48
4.2 手部配件核心计算芯片	49
4.3 其他周边芯片	49
4.4 芯片其他需求和芯片制造	50
4.4.1 交互感知	50
4.4.2 芯片功耗	50
4.4.3 芯片安全	50
4.4.4 芯片制程	51
4.4.5 芯片封装	51
第五章 面向扩展现实的 IP 和芯片平台	52
5.1 现有 IP 如何满足 XR 芯片需求	52
5.1.1 CPU	52
5.1.2 GPU	54
5.1.3 NPU	54
5.1.4 VPU	55
5.1.5 DPU	56
5.1.6 ISP	57
5.1.7 SPU	57
5.2 芯片原型平台	59
5.2.1 一体式 XR 芯片原型平台	59
5.2.2 分体式 XR 芯片原型平台	64
5.3 现有扩展现实设备主控芯片	65

5.3.1 Qualcomm XR/AR 平台	65
5.3.2 Unisoc XR 平台	66
5.3.3 Rockchip XR 平台	66
第六章 扩展现实操作系统	68
6.1 扩展现实操作系统简介	68
6.2 扩展现实核心模块	72
6.2.1 6DoF 头部追踪	72
6.2.2 3/6DoF 手柄	73
6.2.3 Video See-through	73
6.2.4 ATW	75
6.2.5 畸变矫正和色差矫正	76
6.2.6 无线渲染	77
6.3 OpenXR 和 GSXR	79
6.3.1 OpenXR 支持	79
6.3.2 GSXR 支持	81
第七章 未来扩展现实技术展望	83
7.1 无线传输技术	83
7.1.1 Wi-Fi 技术	84
7.1.2 5G/6G 技术	84
7.2 多媒体音视频技术	85
7.3 智能交互技术	86
7.3.1 手势追踪技术	86
7.3.2 眼动追踪技术	88
7.3.3 触觉反馈技术	88
7.4 环境理解技术	89
7.5 生物电感知交互技术	91
第八章 结束语	93

前言

计算机自从诞生以来，不断向着小型化、智能化、个人化方向发展，计算硬件经历了从大型机到小型机、从个人电脑到智能手机的转变，人机交互方式也从早期的打孔纸带发展到键盘、鼠标、触摸屏。人与计算机之间距离越来越近，使用计算机的时间也越来越长，计算机越来越成为人们日常工作和生活中离不开的工具。

以 5G 通信技术和 AI 人工智能技术为引导的新一轮技术革命正在 432 高算力的需求催生出新一代硬件设备平台，以实现真实世界和虚拟世界的进一步融合。扩展现实作为新一轮科技革命的代表技术之一，已被认为是下一代通用性技术平台和互联网的入口，极有潜力改变人与计算机交互的方式，让计算机更好地融入人们的日常工作和生活。

当前扩展现实设备的底层芯片很多是复用智能手机芯片平台的计算能力，扩展现实设备本身的需求与智能手机设备有相通之处，但也有其独特的特点。本白皮书系统介绍了扩展现实设备厂商的需求、技术难点和对底层芯片的需求，详细介绍了扩展现实芯片（以移动芯片平台为主）现状以及未来可能的芯片形态和技术突破点，为从业人员提供一份全面的技术参考，期望能为中国扩展现实芯片产业的未来发展助力。

第一章 扩展现实产业概述

2022年10月28日，工业和信息化部、教育部、文化和旅游部、国家广播电视总局、国家体育总局印发的《虚拟现实与行业应用融合发展行动计划（2022-2026年）》指出：“虚拟现实（含增强现实、混合现实）是新一代信息技术的重要前沿方向，是数字经济的重大前瞻领域，将深刻改变人类的生产生活方式，产业发展战略窗口期已形成。”

一般来说，广义的虚拟现实包括了虚拟现实（Virtual Reality, VR）、增强现实（Augmented Reality, AR）、混合现实（Mixed Reality, MR）等技术及其产业应用，通常用扩展现实技术（eXtended Reality, XR）来指代。

市场规模

根据全球知名数据机构 IDC 数据，XR 全球消费支出在 2022 年已增长至 138 亿美元，并将在 2026 年增长至 509 亿美元，即在未来四年 XR 市场规模将增长 3.68 倍，年度复合增长率高达 32.3%：其中大部分消费市场的增长出现在美国和中国，中国地区占比约为 1/4；消费级和企业级支出的比例大致持平，而 VR 相关支出预计将占总支出的 70%。IDC 认为，XR 市场是“一个准备蓬勃发展的成熟市场”。

此外，根据 Analysis Group 的分析，2021 年兴起后如火如荼的元宇宙产业，如果从 2022 年被采用，并以与移动技术类似的方式发

展，那么预计 10 年后它对全球 GDP 的贡献将达到 2.8%；到 2031 年，它将为全球 GDP 贡献 3.01 万亿美元，其中三分之一(1.04 万亿美元)来自亚太地区。XR 是元宇宙当前最成熟的入口，更是其中最为关键的技术之一，元宇宙也为 XR 产业的长足发展提供了强劲的原动力。由此可见，XR 扩展现实产业的发展对未来全球经济发展至关重要。

行业应用特点

XR 是一个快速发展的领域，与人工智能（AI）和物联网（IoT）被认为是三大未来将显著提高生产率和体验的科学技术。XR 技术可广泛运用于教育、工业、医疗、能源、零售、文旅、展陈、政务、安防等 18 个大类 100 多个应用场景，XR 产业链跨度大，前沿数字技术应用广泛，蕴含巨大发展潜力。

AR/VR/MR 概念界定

XR 指所有由计算机技术和可穿戴设备生成的虚实融合的环境和人机交互，包括 AR、MR、VR 等具有代表性的产品形态。

VR 和 AR 是 XR 概念的主体，两者都可带来虚拟视觉体验，但技术实现路径存在较大的差异：

VR 是利用计算机技术生成模拟环境，强调用户与虚拟世界的实时交互，带来封闭式、沉浸式的虚拟世界体验。VR 主要融合了多模

态感知，包括但不限于：视觉、听觉、触觉、味觉等。当用户处在预定义的区域中时，可以体验到类似真实世界的感官刺激。

AR则是借助计算机图形技术产生物理世界中不存在的虚拟对象，并通过光学透视原理将虚拟对象叠加在物理世界中，带来感知效果更丰富的视觉体验。AR可以将额外的数字视觉信息或者内容叠加到现实环境中，增强虚拟与现实世界的融合与交互。AR具有三个明显特点：数字世界和真实世界相结合、实时自然交互、虚拟对象准确地对应它在现实世界中的实际位置。

MR融合了AR和VR全部要素，将一些虚拟的对象元素添加到真实物理世界中，以便数字对象可以与现实环境交互，从而让用户感觉到这些虚拟元素是真实场景一部分，达到虚实融合的交互体验。当前，部分主流VR设备已可支持MR功能，XR的产品路径和形态正逐步融合迭代。

VR 头显设备形态

VR头显主要可以分成以下三种类型：手机式VR（screenless viewer），主机式VR（Tethered HMD）以及VR一体机（All-in-One）：

1. 手机式VR：手机式VR使用手机的显示屏来进行VR内容的显示，通过USB接口连接手机，代表产品有谷歌的Cardboard以及三星的Gear VR系列。它利用手机处理能力将VR内容进行分屏显示，从而产生立体效果；通过采集IMU（Inertial Measurement Unit）

传感器数据，结合九轴融合算法，计算和预测手机佩戴在头部的实时姿态信息。在移动头部时，眼前显示的内容产生相应变化，可以初步体验到在虚拟现实的情景下观看 YouTube、谷歌街景等。然而，受限于手机性能和软件能力，手机式 VR 产品满足不了用户体验的需求，随着技术发展，逐渐被市场淘汰。当然，Google Cardboard 和三星 Gear VR 对国内市场产生了重大的影响，当年暴风魔镜以及华强北众多的手机式 VR 就是其衍生的现象级产品。

2. 主机式 VR：主要通过有线的方式连接 PC 或者游戏主机等。与手机式 VR 不同，主机式 VR 有独立的光学器件、显示屏以及传感器，从而可以实现 3DoF 或 6DoF 定位。由于 VR 中对算力要求较高的实时定位和渲染模块在 PC 或者游戏主机中完成，主机式 VR 头显对内置的处理器性能要求低，无需集成电池，因此在重量、散热、佩戴等方面非常有优势。但是，有线连接在一定程度上牺牲了设备的移动性，这对用户的沉浸感体验有很大的影响。主机式 VR 头显代表产品有 HTC Vive 系列、Sony PSVR 系列等。
3. VR 一体机。通常认为，VR 一体机是 VR 头戴式设备的终极形态。VR 一体机内置定制的处理芯片、显示屏、传感器和电池，头部跟踪和渲染计算都在自身处理器上运行。此外，它还可以通过无线串流连接到 PC，从而利用 PC 上更丰富的内容及游戏资源。最初的 VR 一体机，比如 Oculus Go，使用了 3DoF 的定位方式，随着处理器性能的提升，6DoF 逐渐成为 VR 一体机的标配。另外，近年来随着 VR 中前置高清摄像头的加入，Video see-through（彩色透视）

功能使得 VR 向着 MR 的方向前进,代表产品包括 Meta Quest Pro, Pico 4 Pro 等。

VR 头显出货预测

根据 Strategy Analytics, Counterpoint 以及 CCS Insights 的预测:到 2025 年,主机式 VR 的年出货量达 1209 万台,VR 一体机可达 3530 万台,年复合增长率分别为 25.5%以及 27.7%。

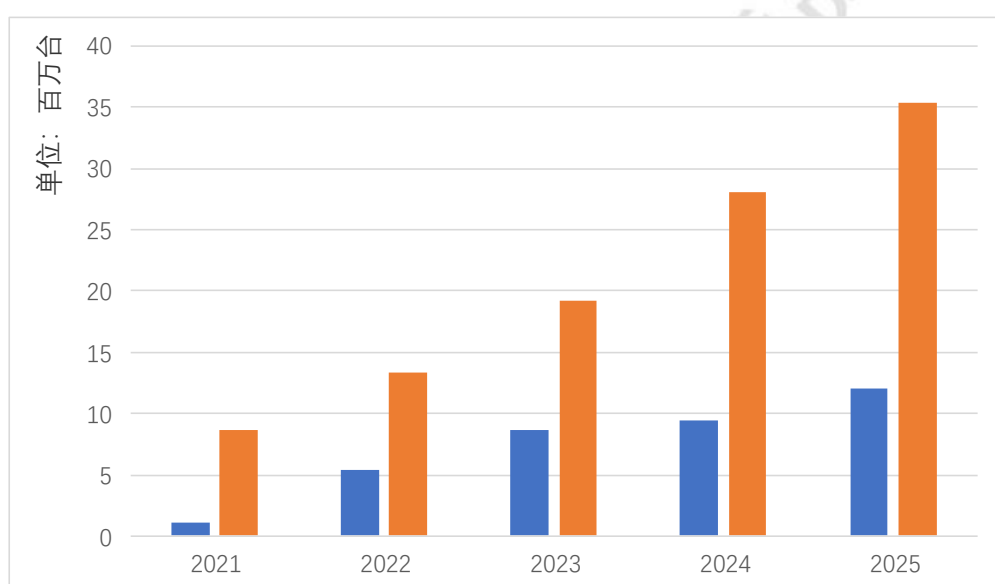


图 1.1 VR 出货量预测(来源: Arm 根据市场信息综合整理)

AR 头显设备形态

AR 头显与 VR 头显有很多相似之处,但是其在便携性、功耗等方面的要求比 VR 更加苛刻。在 VR 中,用户看到的场景和人物全是虚拟的;在 AR 中,用户看到的场景和人物一部分是真实的,一部分是虚拟的,需要把虚拟的信息带入到现实世界中,这种显示方式对计算机视觉处理要求更高。目前的设备受限于处理器的能耗比,还无法做到在期望的小尺寸整机上实现高性能的算力。

AR 头显可以分成以下两种类型：一体式 AR 和分体式 AR。

1. 一体式 AR：计算和显示单元融为一体。一体式 AR 眼镜分为两类，一类是高算力一体机，一类是轻量级一体机。高算力一体机本身可独立完成计算和内容显示。轻量级一体机本身是一个可独立运行的硬件，但显示内容一般需要外部通过无线传输到眼镜。虽然目前的芯片和光学技术还很难在眼镜形态下兼顾性能和功耗，但是芯片厂商、供应链企业以及 AR 终端产品公司都在发力，AR 一体机成为消费级产品指日可待。
2. 分体式 AR：显示单元和计算单元分离。显示单元运行一套轻量级 Linux 或 RTOS（实时操作系统），主要负责传感器数据的采集、传输以及显示。计算单元（包括头部 3DoF/6DoF 跟踪、手势跟踪、眼动跟踪、实时渲染等部分）通常在外部的计算单元中实现，比如手机、外置计算盒子等。与 VR 相比，AR 眼镜对重量尺寸、交互定位、佩戴体验有更高要求，所以 AR 眼镜的发展要迟于 VR 头显，但是随着 birdbath、自由曲面等光学显示技术的成熟和成本的下降，AR 的增长速度已经明显超越 VR，行业预计将在未来 3-5 年迎来爆发期。

AR 头显出货预测

根据 Strategy Analytics, Counterpoint 以及 CCS Insights 的预测，到 2025 年，分体式 AR 头显的年出货量可达 2900 万台，一体式 AR 头显可达 290 万台。

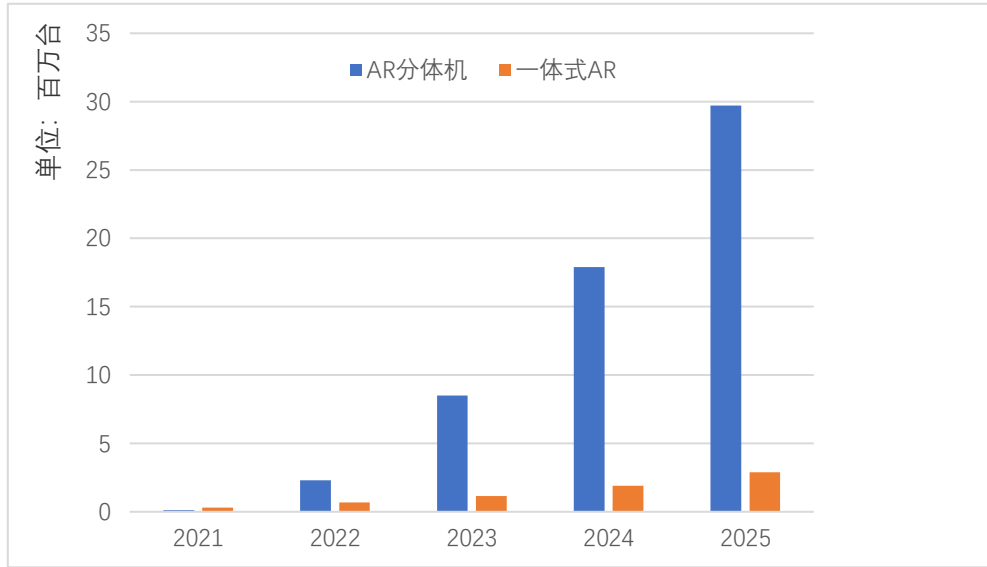


图 1.2 AR 出货量预测(来源: Arm 根据市场信息综合整理)

扩展现实产业及标准推进
(TC625)

第二章 现有热门扩展现实设备方案概览

伴随近几年各大厂商对扩展现实设备的探索，技术要求逐渐成型，出现分体机、一体机等多种样式，以适应不同场景与需求。以下为主流设备相关参数与部分产品介绍：

2.1 VR 一体机概览

VR 一体机在性能与画面要求上高于手机。芯片与系统基于手机移动生态演进，在 Arm 架构和 Android 系统的组合上进一步发挥，需求上也与手机类似，更加关注性能、功耗、重量间的平衡。芯片层面，主芯片目前以高通骁龙 XR2 为主，与 Arm[®] Cortex[®]-A77 + A55 性能相当，其功耗控制在 8 瓦左右，电池 3500-5000 毫安时。系统普遍基于 Android 深度定制，可在一定程度上沿袭 Android 开发者与生态。同时，行业也开始出现在手柄中采用中低端 SoC 的探索，如 Quest Pro 手柄中使用高通骁龙 662 芯片，用于处理手柄上的 3 枚摄像头实时数据。视觉层面，分辨率普遍接近单眼 2K，刷新率 90Hz-120Hz，并通过超短焦折叠光学 (Pancake) 逐渐取代菲涅尔光学 (Fresnel) 以进一步降低机器的厚度与重量，视场角为 90-100° 左右，并开始采用彩色透视。功能层面，VR 一体机普遍支持 6DoF 头部跟踪与 6DoF 手柄交互，并在旗舰设备中加入手势跟踪、眼球跟踪与面部识别。

2.1.1 VR 一体机热门设备参数

VR 一体机	Pico4	Quest2	Quest Pro	Skyworth	YVR 2
--------	-------	--------	-----------	----------	-------

					Pancake1	
基本信息	发布时间	2022-09	2020-10	2022-10	2022-07	2022-07
	主芯片	高通骁龙 XR2	高通骁龙 XR2	高通骁龙 XR2	高通骁龙 XR2	高通骁龙 XR2
	手柄芯片	/	/	高通骁龙 662	/	/
	操作系统	Pico OS 5.0 (Android)	Android	Android	Android	Android
视觉	单眼分辨率	2160x2160	1832x1920	1800x1920	2280x2280	1600x1600
	刷新率	90Hz	90Hz	90Hz	90Hz	90Hz
	显示	Pancake	菲涅尔	Pancake	Pancake	Pancake
	视场角	105°	97°	106°	95°	95°
	透视	彩色	黑白	彩色	彩色	黑白
功能	体感追踪	6DoF	6DoF	6DoF	6DoF	6DoF
	手势识别	支持	支持	支持	/	/
	眼球识别	/	/	支持	/	/
	面部识别	/	/	支持	/	/
其他	电池容量	5300mAh	3640mAh	5000mAh	5500mAh	5300mAh
	整机重量	596 克	503 克	722 克	436 克	630 克

表 2.1 VR 一体机图表 （来源：VR-Compare，公开信息整理）

2.1.2 VR 一体机国内设备介绍

PICO

PICO4 为字节跳动旗下 VR 品牌，并于 2022 年发布 PICO4。PICO4 为用户提供了丰富的内容，包括 VR 运动健身、VR 视频、VR 娱乐和 VR 创造四大场景。



PICO4 全系搭载高通骁龙 XR2 处理器，采用 Pancake 折叠光路设计，让设备更轻，采用前后平衡分布的设计以提升佩戴舒适性，支持 4K+ 超视感屏和 105° 超大视角，采用 HyperSense 振感手柄提高振感反馈，通过 6DoF 空间头手定位让设备精准识别，支持裸手交互，更具沉浸感。

创维

创维 VR 旗下的 PANCAKE 1 是消费级短焦 VR 一体机，具有超薄机身、自研的 6DoF 追踪定位和高通骁龙 XR2 的顶级配置。创维 PANCAKE 1 作



为消费级短焦 VR 一体机，采用 Pancake 超短焦折叠光学技术，主机仅 189g，整机仅 436 克，相比传统菲涅尔镜光学体积缩小了 3/4，机身厚度仅 32mm，佩戴舒适度大为升级。轻薄小巧的主机支撑实现 0-90 度的上翻，用户可以随时在虚拟和现实世界之间切换，并且支持 0-500 度近视屈光度调节，59mm-68mm 无极瞳距调节。创维 PANCAKE 1 采用

自研的 6DoF 追踪定位，头部、手部的双 6DoF 空间体感定位，定位精度可达毫米级，配备存储为 8G+128G。

YVR

上海鱼微阿科技有限公司（简称“YVR”）于 2022 年推出 Pancake 超短焦折叠光学 VR 一体机 YVR2，是国内唯一同时斩获德国红点和 IF 两大全球工业设计顶级大奖的 VR 眼镜。



“YVR2”采用自研 VSLM 算法，实现业内领先的头手 6DoF 空间定位，并具有多地图记忆功能。YVR 产品上线了丰富的 C 端应用，如观影、图书馆、书法等休闲娱乐类应用，以及乒乓球、篮球、射箭、攀岩、拳击等运动类应用，给用户带来全新的沉浸式三维空间体验。

2.2 主机式 VR 概览

主机式 VR 多数依赖于 PC/主机进行计算，相比于 VR 一体机，可以放开对于性能与续航的限制，更加注重提升视觉、操作的体验。芯片层面，由于主机式 VR 依赖外部连接的 PC/主机，并未在设备中集成高性能 SoC，所依赖的操作系统也多以 Windows 等为主。视觉层面，主机式 VR 的分辨率和刷新率由于脱离了功耗限制，普遍高于 VR 一体机。功能层面，以支持 6DoF 为主，手势识别/眼球识别/面部识别等功能，需要增加额外的功能模块。

2.2.1 主机式 VR 热门设备参数

主机式 VR		HTC Vive Pro2	Sony PS VR2	Pimax 5K Super	DPVR E4
基本信息	发布时间	2021-06	2023-02	2020-12	2023-02
	操作系统	连接 PC Windows	连接 PlayStation	连接 PC Windows	连接 PC Windows
视觉	单眼分辨率	2448x2448	2000x2040	2560x1440	1832x1920
	刷新率	120Hz	120Hz	180Hz	120Hz
	显示	菲涅尔	菲涅尔	菲涅尔	菲涅尔
	视场角	120°	110°	200°	116°
功能	体感追踪	6DoF	6DoF	6DoF	6DoF
	手势识别	/	/	/	/
	眼球识别	支持（扩展配件）	支持	/	/
	面部识别	支持（扩展配件）	/	/	/
	重量	800 克	560 克	750 克	495 克

表 2.2 主机式 VR 图表（来源：VR-Compare，公开信息整理）

2.2.2 主机式 VR 国内设备介绍

Pimax

Pimax Vison 5K Super 的 VR 头显，是业内第一款达到 180Hz 刷新率的消费级 VR



头显。该产品采用两块分辨率为 2560×1440 的 LCD 屏幕，最大视场角达 200 度，刷新率则提升到 180Hz。不过，由于带宽的限制，当将 Pimax 5K Super VR 头显设置为 160Hz 或更高刷新率时，视场角将降低至 170° 。此外，Vision 5K Super 的瞳距调节范围为 55mm-75mm，覆盖大部分人群，并且鼻部有防漏光设计保障外部光线不会干扰 VR 体验，基于人体工学设计并希望拥有良好的佩戴体验。

DPVR



DPVR E4 支持用户自定义 VR 头显外观，并配备两个控制器，采用 90° 上翻设计。DPVR E4 液晶显示屏支持 120Hz 刷新率、4K 分辨率，配备四颗摄像头，可实现无基站定位；支持 6DoF，可全面检测头部的空间与角度信息；使用菲涅尔透镜，视场角（FOV）可达 116° 。

2.3 AR 一体机概览

AR 一体机相比于 VR 头显设备，需佩戴更长的时间，使用场合也不局限于固定的室内等场所，所以对于重量问题最为敏感。为了控制重量需要做出更多取舍，导致功能与形态最多样，AR 一体机可大致分为以下两类：

- 1、高算力一体机：主要用在工业、商业场景，这些场景为了完整功能与体验，往往体积大、重量高。芯片层面，工业、商业场景对于功

能要求较多且复杂，主芯片配置与高端移动芯片相当。视觉层面，设备单眼分辨率以 1.5K 为主。功能层面，多数支持 6DoF、手势识别和眼球识别。

2、轻量级一体机：主要用在消费场景，这些场景往往精简功能，关注细分使用需求。芯片层面，多采用智能手表芯片，系统多数以 Android 为主。视觉层面，单眼分辨率普遍是 VGA (640x480) 水平。功能层面，消费场景较为多样，信息提示、翻译、摄像等各有不同，主要依靠无线传输从其他设备获得。

2.3.1 AR 一体机热门设备参数

AR 一体机		大算力 一体机		轻量级 一体机			
		Microsoft HoloLens2 (双目)	Rokid X-Craft (双目)	OPPO Air Glass 2 (双目)	OPPO Air Glass 1 (单目)	Xiaomi Mijia (单目)	影目 INMO Air (单目)
基本信息	发布时间	2019-11	2020-11	2022-12	2021-12	2022-08	2022-06
	主芯片	高通骁龙 850	Amlogic A311D	Apollo4	高通骁龙 Wear4100	高通骁龙	Arm® Cortex®-A53
	操作系统	Windows 10 HoloGraphic	YodaOS-XR	RTOS	Android	Android	IMOS (Android)
视觉	单眼分辨率	1400x936	1280x720	640x480	640x480	/	640x400
	刷新率	60Hz	60Hz	60Hz	60Hz	/	/

	视场角	52°	40°	28°	28°	/	26°
功能	体感追踪	6DoF	6DoF	/	/	/	3DoF
	手势识别	支持	支持	/	/	/	/
	眼球识别	支持	/	/	/	/	/
其他	电池容量	4600mAh	10000mAh	150mAh	246mAh	1020mAh	350mAh
	重量	566 克	780 克	38 克	30 克	100 克	78 克

表 2.3 AR 一体机图表（来源：VR-Compare，公开信息整理）

2.3.2 AR 一体机国内设备介绍

OPPO



2021 年 12 月,OPPO 发布了 OPPO Air Glass 1, 此款眼镜采用单目单色显示, 其创新性获得由工信部

主办的 2022 世界 VR 产业大会“VR/AR 创新金奖”, 2023 年荣获国际权威奖项“2023 爱迪生奖银奖”。



2022 年底, OPPO 发布了第二代的 OPPO Air Glass2, 整体重量仅 38g, 采用全新双目设计, 自主光机架构设计。OPPO 研发了全球首款树脂衍射光波导镜片, 并

专注于光引擎小型化, 微型光机相较上一代减小 40%。此外, Air Glass

2 还能进行许多实用功能使用，例如：信息提示、地图导航、语音翻译、提词器等，语音转文字的功能则能够帮助听障人士进行日常交流。

Rokid

2020 年 11 月，Rokid 发布一体式工业 AR 头环 X-Craft，作为一款国产化 AR 设备，X-Craft 是全球首款搭载 5G 模组的工业三防、本安一区防爆 AR 头环，可与标准安全帽完



美适配，具备 IP66 防护等级，头环采用双目衍射光波导和 DLP 光机近眼显示技术，视场角 40 度，入眼亮度默认配置 400nits，透光率超过 80%，一体式设计，集成了电池、5G、Wi-Fi6、蓝牙 5.0，支持离线语音、实体旋钮、手势等多种交互模式，支持 GPS、北斗等定位，拥有 4G 内存和 128G 闪存配置，可满足更多工业企业的应用场景。

2.4 AR 分体机概览

AR 分体机依赖外接手机，不同产品间形态较为接近，并开始出现外接计算单元形式解决手机兼容性问题。主芯片层面，由于 AR 分体机通过连接手机等外设进行计算，眼镜本身不集成高性能 SoC，外接计算单元的 SoC 性能与机顶盒芯片相当，系统也使用 Android，操作体验类似机顶盒系统。视觉层面，单眼分辨率以 1080p 为主，刷新

率 75Hz/90Hz，视场角在 45° 左右。功能层面，目前主要以观影为主，主要支持 3DoF。

2.4.1 AR 分体机热门设备参数

AR 分体机		Rokid Air	Nreal Air	OPPO AR Glass 2021	雷鸟 Air 1S	Rokid Max
基本信息	发布时间	2021-09	2022-08	2020-12	2022-10	2023-03
	主芯片	/	/	/	/	/
	操作系统	连接 Rokid Station	连接手机	连接手机	连接手机	连接 Rokid Station
光学显示	单眼分辨率	1920 x1080	1920 x1080	1920 x1080	1920 x1080	1920 x1200
	刷新率	75Hz	90Hz	60Hz	60Hz	120Hz
	视场角 (FOV)	43°	46°	43°	45°	50°
功能	体感追踪	3DoF	3DoF	6DoF	3DoF	3DoF
其他	电池容量	Rokid Station 5000mAh	/	/	/	Rokid Station 5000mAh
	重量	83 克 (含鼻托)	79 克 (不含鼻托)	117 克 (含鼻托)	82 克	75 克 (含鼻托)

表 2.4 AR 分体机图表 (来源: 公开信息整理)

2.4.2 AR 分体机国内产品介绍

Rokid

2021年9月，Rokid 第一款消费级 AR 智能眼镜 Rokid Air 发布，拥有 83g 超轻量化，提供 43 度广域视角体验，及 500 度以内的屈光度调节。



2022年10月，Rokid 首创推出 Station 的形式代替手机，连接 Rokid Air 使用，解决行业手机兼容问题的同时，也让 AR 眼镜首次成为独立的生态。2023年3月，Rokid Max 发布，自重 75g，拥有约 50 度 FOV，实现 6 米 215 寸屏幕，最高支持 120Hz 超高刷新率，入眼亮度最高 600nits，支持 600 度以内屈光度无损调节。Rokid 致力于 AR 眼镜等软硬件产品的研发及以 YodaOS-XR 操作系统为载体的生态构建，通过操作系统、语音技术、计算机视觉、光学显示、芯片平台、硬件设计等多领域研究，基于 AI+AR 技术，在消费、数字文化、医疗、工业制造、汽车和企事业等领域打造了多款 AR 产品和解决方案，尤其在数字文化领域，已在全球百余家博物馆和景区落地。

Nreal

Nreal Air 是 Nreal 打造的一款 AR 眼镜，拥有空间视网膜级显



示，配备 130 英寸空中投屏和 201 英寸 AR 锐彩天幕，并支持与手机、平板、笔记本、掌机、游戏主机等多种设备的连接。Nreal Air 搭载自主研发的“惊鸿锐影”光学引擎 2.0，双目分辨率 3840*1080，MTF>0.7，PPD=49。同时，Nreal Air 拥有 100000:1 的超高对比度、108%sRGB 高色域，入眼亮度高达 400nits。Nreal Air 是全球首款获得 TUV 莱茵低蓝光、无频闪、Eye comfort 三项权威认证的 AR 眼镜。

OPPO



OPPO AR Glass 2021 于 2020 年 11 月发布，采用有线分体式的产品形态，使用鼻梁和耳朵的支撑方式，通过 USB 线连接手机。显示方案使用 Micro OLED 搭配 Birdbath，显示分辨率 1920×1080，FOV 43°。双目 SLAM，双目 Camera 分辨率 640×400，FOV 140°。使用 TOF 实现主动式三维建模，RGB camera 支持 500W 像素，支持无线投屏。

雷鸟创新

由 TCL 电子孵化的雷鸟创新在 Birdbath 和光波导领域均已具备产品化能力，是行业技术布局完整的 AR 品牌。2023 年初，



雷鸟创新在美国 CES 上发布了消费级真 AR 眼镜雷鸟 X2，在 MicroLED 全彩显示、自研光引擎精密及小型化等关键技术上取得突破性进展。该款眼镜搭载高通骁龙 XR2 专业算力芯片平台，配备 1600 万像素高清摄像头，可为用户提供沉浸式英语学习、智能翻译、实时导航、信息提醒、快捷拍照等场景服务，打造便捷高效的 AR 体验。

扩展现实产业及标准推进委员会
(TC625)

第三章 扩展现实设备底层技术需求

3.1 XR 基本组件

虽然 XR 设备形态各异，但其基本组件都类似，都有感知单元、计算单元、显示单元以及其他组件，只是在不同应用场景中，受制于重量功耗等硬约束条件，对各组件的功能和性能各有取舍。随着技术的发展和算力的提高，VR 和 AR 最终可能殊途同归，从差异化产品形态最终走向融合。

感知单元

目前大众熟知的手机、电脑等设备的信息输入以触摸屏、键盘为主，XR 设备则需要大量依赖传感器，如通过摄像头、IMU 传感器（陀螺仪、加速度计等）配合 SLAM 算法进行实时空间定位计算，跟踪用户头部运动，部分设备支持手势跟踪、眼球跟踪和面部识别。相比于手机，XR 设备需要更多的摄像头与传感器。同时，手柄目前承担着用户操作输入的主要功能，多数手柄配备陀螺仪、加速度计等传感器，Meta Quest Pro 则直接在手柄中配备高通骁龙 662 SoC 与多个摄像头，进行 SLAM 计算，就近完成计算工作，同时降低主 SoC 负担。

不同类型的传感器响应来自真实环境的物理或化学刺激，并为系统的开发提供必要的的数据。XR 所需的常见传感器包括：

- 1) 深度传感器：测量深度和距离；

-
- 2) 陀螺仪：检测运动角速度的变化；
 - 3) 距离传感器：测量物体距离；
 - 4) 加速度计：检测运动加速度的变化；
 - 5) 光传感器：测量光强度和亮度。

计算模块：

计算模块属于 XR 设备中占整机成本最高、计算任务最重的模块。

从功能与组成角度，VR 设备的主板与移动主板设备类似，包含 SoC、内存、存储等计算设备所需的基础组件，其中 VR 一体机由于需要在机身内承担所有的计算、渲染等任务，往往采用与旗舰手机相当的 SoC，部分设备采用 CPU 和 RAM 堆叠的形式进一步压缩主板空间。以 VR 一体机主流的 SoC 高通骁龙 XR2 为例，采用 4 大核+4 小核的设计，通过大核满足高性能场景，通过小核兼顾日常应用与续航，同时支持多路摄像头并行处理，专用 DSP 视觉计算等功能。而 VR 分体机由于依赖 PC、游戏主机进行计算，主板不需要配备大算力 SoC，只需要进行视频处理与显示、音频、传感器数据采集与传输等工作。

AR 设备的发展从体感上来看，正在从笨重、形状突出的“巨型”设备逐渐向生活中日常佩戴的眼镜形态转化。由于轻量级 AR 眼镜对尺寸、功耗和散热有更高要求，一体化的计算芯片单元也正在逐渐拆分成多个封装、更小的协处理器单元，从而可以分散到眼镜的不同部

位来缩减结构尺寸和提高散热性能。目前一种可能的方向是将 AR 的计算单元拆分为三个主要组件，分别为 AR 处理器，AR 协处理器以及连接单元。AR 处理器主要负责感知以及显示的输出，一般支持多路摄像头的并行处理，进行 6DoF 的运动跟踪和定位，提供低延迟下对于用户动作的快速反应。AR 协处理器一般聚集了各种传感器，来完成一些 AI 以及计算机视觉的工作。AR 协处理器结合了摄像头和传感器数据，在现在的设备中往往还会含括眼动追踪的模组进行注视点渲染，来降低算力的需求。最后的连接单元主要搭载低延迟、低功耗的 Wi-Fi6/6E/7 模块，提供低延迟下的高吞吐量，确保 AR 眼镜能以极低的延迟将传感器数据传给手机或 PC，并从手机或 PC 获取计算和渲染好的内容以显示。随着摩尔定律的演进，未来如果有单一的一体化芯片能实现高算力、低功耗，也可能各部分重新融合，在眼镜上完成全部运算。

显示模块：屏幕与镜片

XR 设备所采用的“光学方案与微显示屏”是影响其视场角、透光度、亮度、刷新率、画面显示效果等显示功能的关键因素之一。VR 设备利用光学透镜将由靠近人眼的微显示器屏幕上所显示的图像聚焦、放大、投远，投射到人眼中，佩戴眼镜时所看到的图像完全由显示屏提供，实际场景的自然反射光线无法通过眼镜直接进入人眼，佩戴者通过光学和微显示屏获得了完全虚拟、沉浸的视觉内容体验。AR（增强现实）一般是光学镜片将位于人眼侧的显示器屏幕上的图像通

过光学镜片耦入、耦出、扩瞳，以实现聚焦、放大、投远，投射到人眼中，同时真实世界的反射光线也能通过镜片进入佩戴者眼中，佩戴者通过光学和微显示屏获得了虚实融合的视觉内容体验。

VR 设备的显示模块与移动终端、桌面电脑等设备存在较大的差异，在大众较为熟知的 LED/LCD 屏幕作为基本显示功能的基础上，再通过叠加光学镜片，使得人眼能够近距离观看，并在左右眼通过不同的屏幕与镜片显示画面产生 3D 视差，实现 3D 显示效果的目的。为了实现立体效果，需要对双目所见图像单独渲染，这大大增加了图像渲染的工作量。要实现逼真的显示效果，对图像分辨率和刷新率也有较高要求，高分辨率和高刷新率都会带来对数据高算力及传输高带宽的要求。

基础模块：电源管理、网络连接、音频等

智能设备所需要的各种基础功能，在 XR 设备中并无太大差别，比如通过电源管理芯片进行 LED 驱动、负荷开关、充放管理、过电压保护等功能，通过 Wi-Fi、蓝牙等芯片，实现主流的连接与数据传输，以及目前的基础音频播放功能和未来需要更多算力的空间音频等功能。

3.2 主要参数指标要求

无论是消费类或是垂直行业，都需要关注几个设备的基本参数指标来达到一定的沉浸效果。应用领域的差异会导致这些基本参数的需

求各有不同，设备厂商会根据场景的需要做硬件方面的调校及平衡，各参数对于用户体验的影响有所不同。以下是根据参数的重要性所列出的基本设备指标。

3.2.1 延迟

VR 延迟率指 VR 头盔显示设备的视觉观测与头部运动的匹配程度，又称为运动到成像时延（Motion-to-Photon Latency），也就是当新的图像被渲染后以及主设备捕捉到各类传感器而产生反馈传递给用户显示的时延。主流的应用场景中，观影、游戏及教育培训都对设备的画面的延迟有着极高的要求，延迟的大小可以给用户带来截然不同的体验。延迟的优化可以让设备显示与大脑预期相匹配，将人体的眩晕感降到最低，提高沉浸感，所以延迟指标是 XR 设备的重中之重。VR 为了实现沉浸效果，对用户体验到的 MTP 延迟建议在 20ms 之内，经过算法预测的 MTP 可以降低到 10ms 以内。影响延迟的因素有很多，主要有屏幕的刷新率、GPU 的渲染速度、无线/有线的稳定传输、编解码技术、6DoF 算法、异步时间/空间扭曲技术（ATW/ASW）等。

3.2.2 视场角

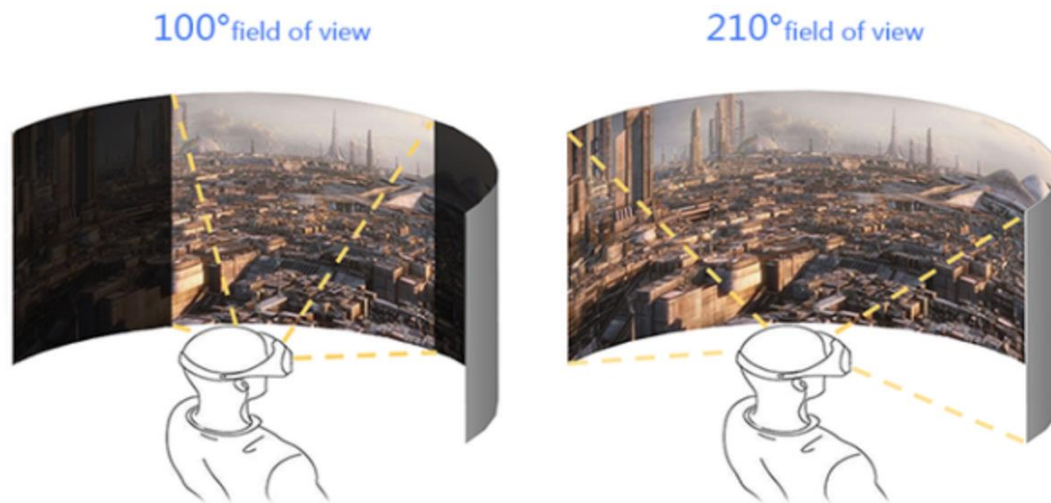


图 3.1 视场角示意图

视场角 (Field of View, FOV) 是人眼通过光学设备看到的一个开放的、可捕获的最大区域，换句话说就是扩展现实设备允许用户看到的最大画面。FOV 通过水平、垂直以及对角线来评估，FOV 越大用户看到的视场越大，沉浸感越强。FOV 在 XR 应用中扮演着重要的角色。目前 VR 的 FOV 可以做到 120 度，AR 的 FOV 在 30-70 度之间。通常情况下，人眼能看到的 FOV 是水平 220 度、垂直 120 度。XR 设备 FOV 与人眼的 FOV 差距来自于光学技术、显示技术、芯片技术等约束。

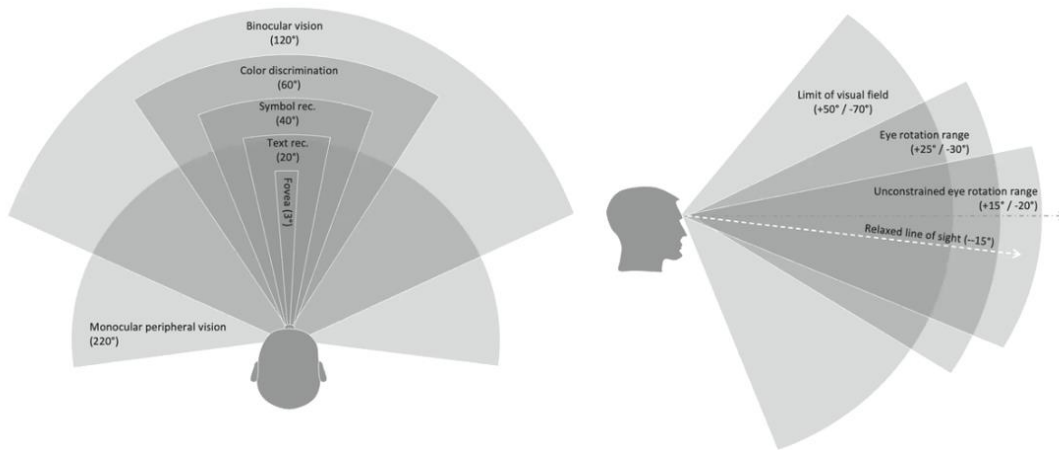


图 3.2 人眼视场 (FOV) 范围 (来源: Digital optical elements and technologies , 2019)

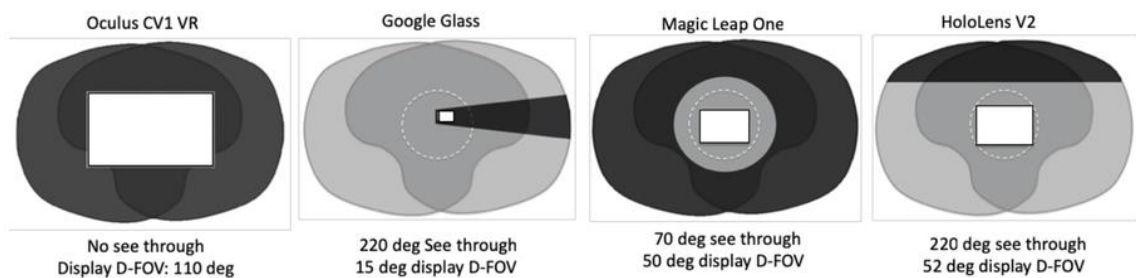


图 3.3 XR 设备 FOV (白框范围内) vs 人眼的视场 (来源: Digital optical elements and technologies , 2019)

3.2.3 角分辨率

不同于手机屏幕用 PPI (Pixel per inch, 每英寸像素) 来判定屏幕清晰度, 头显产品的清晰度是通过 PPD (Pixels Per Degree, 角分辨率) 这个参数来衡量。因为头显是透过光学系统看屏幕放大的虚拟图像, 而不是直接看屏幕, 单用 PPI 是无法衡量头显的清晰度。角分辨率描述了在每个视觉度中实际看到的像素数——每度像素数, 并解释了头戴式设备之间的视野差异。角分辨率的计算方式是单目屏

幕分辨率/视场角，人眼正常视力下的分辨能力为 50-60PPD。一般，在观看动态视频的时候，只要达到 30PPD 左右，人眼就感觉不到明显的纱窗效应，基本满足用户体验。目前主流 AR 设备的 PPD 接近 50，主流 VR 设备的 PPD 在 20 左右。

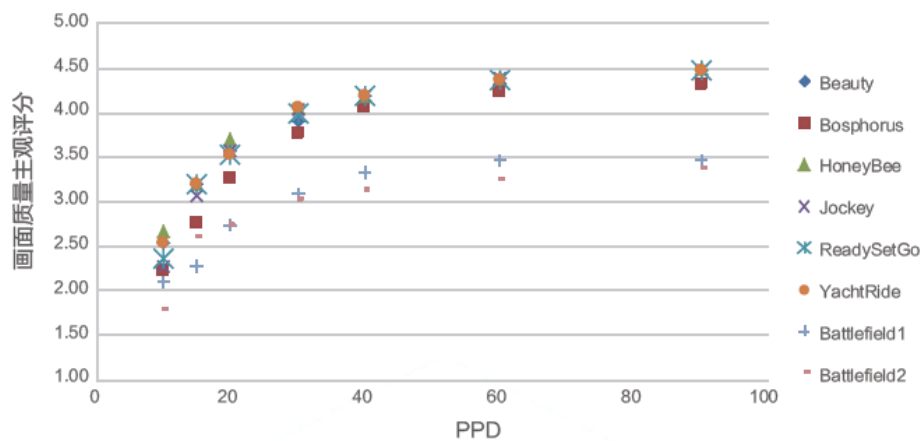


图 3.4 PPD 与画面质量的关系（来源：VR 体验标准技术白皮书）

早期的 VR 设备由于屏幕和内容的分辨率不足，人眼会直接看到显示屏的像素点，这被称为“纱窗效应”。2019 年至今随着一系列软硬件技术的迭代，屏幕的分辨率已经提升到了 4K（单眼 2K），甚至有的已经达到单眼 4K 的效果，极大地提升了画面清晰度，这也是近两年扩展现实又迎来一轮新高潮的关键因素之一。

当前的单眼 2K 离完美的视觉体验依然有很大的提升空间，单眼 4K 以及 90Hz 或 120Hz 的刷新率可以让视觉体验再上一个台阶。伴随着高规格的显示屏幕，更高的带宽以及更高的音视频处理能力，硬件设备也需要一定的升级，所有的升级会造成整机的成本至少翻 2 到 3

倍。当前市场上大众的消费能力以及内容供应还无法匹配整机成本的变化，但随着产品迭代、技术发展和成本的降低，用户会逐步获得更高清的元宇宙体验。

3.2.4 续航及重量

XR 高水平的硬件性能需求，使得渲染、计算、传输、感知等方面的功耗增加，综合下来比传统的智能终端的功耗更高。同时在体验特点方面，为保证 XR 设备的舒适度，不仅需要 XR 设备轻量化，而且需要强大的续航能力。XR 设备物理形态的变化、外接电源、外接电池等手段，可以用来提高续航时间。目前主流 XR 设备的续航时间为 2~3 小时，续航时间短和设备重量大已经成为阻碍 XR 未来发展的重要因素。国内外厂商也在积极的解决这一问题，例如设备配件厂商 Rebuff Reality 推出的 Quest 专用电池模块 VR Power2 配备 1 万毫安时电池容量，续航时间可延长到 8 小时以上。

光学模块、电池以及散热组件成为 XR 设备的主要重量来源，目前主流的 VR 头显重量在 500g 左右。未来随着先进的芯片制程和设计开发、高光效低功耗的光学解决方案、分布式计算架构等领域的不断演进，有望进一步降低 XR 设备功耗。在人体工学设计，电池技术的研究突破有助于实现 XR 设备轻量化设计。

3.3 核心算法技术

3.3.1 DoF 技术

DoF 指“自由度”。3DoF 指的是 3 个旋转运动：允许用户头部处于一个固定位置时，可以前后翻转、左右摇摆、水平转动。6DoF 相比 3DoF 包括额外的 3 个轴的位移运动，通过相机与陀螺仪、加速度计等传感器结合 SLAM 算法，允许用户身体在 X、Y、Z 三个轴的移动。

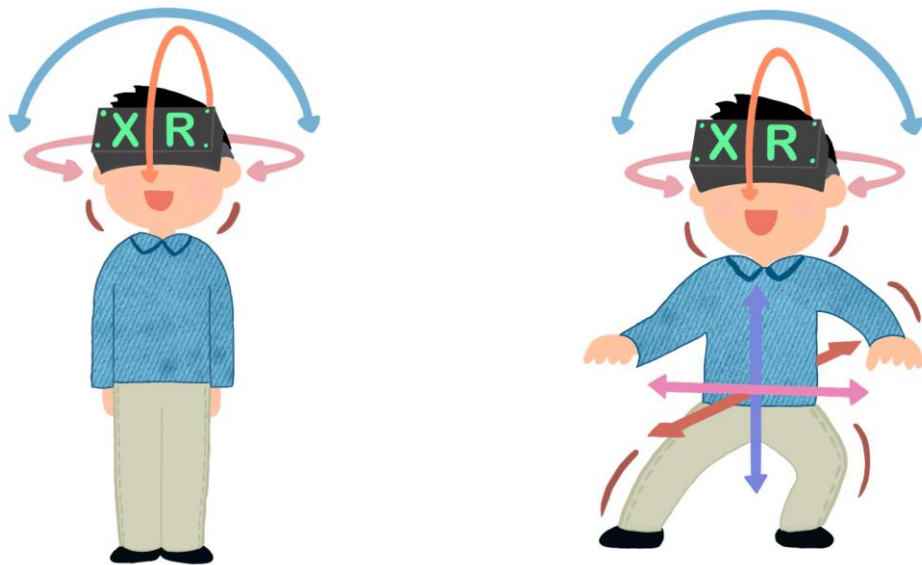


图 3.5 XR 设备 3DoF 模式（左图）和 6DoF 模式（右图）

3DoF 的设备一般用来看电影、看直播，而当身体需要移动并且需要与虚拟或现实场景产生交互时，6DoF 便是必须的。6DoF 头显设备有两种定位方式，一种是固定位置标记（Outside-in），另一种是使用计算机视觉算法从集成摄像头计算位置（Inside-out）。市场上，早先出现的是 Outside in 方案，又称为由外而内追踪方案，需要依靠事先放置的定位装置跟踪 XR 设备上固定标记，并要求设备在定位

点覆盖范围内才能计算位置。而另一种是 Inside-out 方案，又称为由内而外跟踪方案，它不需要架设额外的定位装置，依靠 XR 设备的摄像头，让设备自己检测外部环境变化，经过视觉算法（SLAM 算法）计算出摄像头的空间位置即可。3DoF 设备的最大缺点是它们不太适合沉浸式内容，用户将很难与虚拟环境进行无缝交互，这不仅会降低沉浸感和使用乐趣，还有可能引发晕动症。3DoF 设备的总体趋势正在逐步下降，越来越多的厂商将支持更高交互能力的 6DoF 设备。



Inside-out



Outside-in

图 3.6 Inside-out 和 Outside-in 示意图

3.3.2 ATW/ASW 技术

XR 用户在使用设备的过程中，头部经常处于持续转动的状态，当一帧图像在 GPU 渲染好后，输出到显示之前，用户的头部和视线可能已经发生了一定角度的转动，这会导致用户看到图像和现在所处的位置应该看到的图像有所偏移，如果渲染时间太长会发生丢帧现象，画面会产生颤动和延迟从而导致体感眩晕，因此 Oculus 推出了“异步时间扭曲（Asynchronous Timewarp）”技术，简称 ATW。ATW 会在

画面渲染的同时追踪用户的头部动作，它采用与几何扭曲相同的概念，并用它来补偿丢帧。如果当前帧没有及时完成渲染，ATW 将使用最新的位置数据重新投影前一帧，经过矫正生成中间帧。之所以称为“异步”是因为它与渲染并行发生，而不是在渲染之后发生。在知道真实帧是否会按时完成渲染之前，合成帧已经准备就绪。ATW 算法需要强大的 GPU 硬件支持，实时性要求很高，并且非常依赖 GPU 多线程和细颗粒度抢占能力支持。ATW 只能解决用户头部转动（3DoF）带来的画面问题，而不能覆盖身体的平移或场景中其他物体的运动（6DoF）。

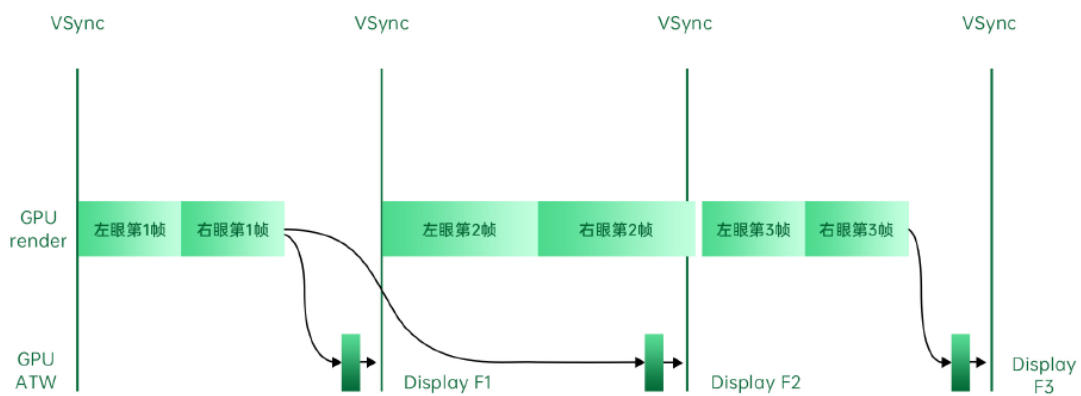


图 3.7 ATW 示意图

2016 年 12 月，Oculus 发布了异步空间扭曲（ASW）技术来解决 ATW 无法解决的运动平滑问题。ASW 本质上是一种快速外推法，它使用先前帧之间的差异来预测下一帧应该是什么样子。ASW 并不是长时间启用的，当应用程序在几秒钟内持续丢掉多个帧时，ASW 会自动启用，然后它会强制应用以半帧速率（45FPS）运行，并每隔一帧合成

生成一次。因此 ASW 不会取代 ATW，ATW 始终处于激活状态而 ASW 只在需要的时候启动。当前的 ATW 和 ASW 都是软件方案，未来将有机会升级为硬件方案，减少系统资源占用并且极大程度上提升效率。通过 ATW 和 ASW 的技术可以使设备在帧率不够或者 GPU 渲染跟不上时，让画面更加流畅，从而给用户带来更顺滑的体验。

除了现在主流的 ATW/ASW 方案，还在不断的产生新的算法方案来补偿画面以及帧率的缺陷。通过 AI 深度学习算法来自动补偿丢失帧或者来预测下一帧也是该领域的研究方向之一。与传统的 ATW 和 ASW 不同的是，深度学习的方案将更加依赖于神经网络芯片而不是 GPU，这样的方案可以极大程度上降低 GPU 的高负载，同时降低功耗。介于现阶段技术及市场的成熟性，还没有专门用于优化帧率的 DPU 或者 AI 芯片出现在当前主流的 VR 硬件平台里，但随着技术的一步步完善，相信更加多样性的算法方案以及芯片方案会顺势而生。

3.3.3 眼动追踪

眼动追踪日渐成为 XR 中高端产品的标配，代表性的产品有 Meta Quest Pro 和微软 Hololens2，且应用场景多元化。随着技术突破和用户需求的提升，眼动追踪给用户体验带来的价值逐渐落地在 XR 产品上，比如，注视点追踪、用户瞳距适配、生物识别、疲劳状态监控等。注视点追踪可以用在眼控交互、可变注视点渲染、可变焦显示系统中的辐辏调节冲突控制等场景中。眼动追踪技术主要有基于特征和基于图像两种方案，两种方案都需要红外摄像头和 LED 硬件实现。目

前市场上大部分 XR 产品采用的是基于特征方式做眼动追踪，通过光线在角膜外表面上的斑反射来推算瞳孔位置。

3.3.4 注视点渲染

XR 设备所生成的虚拟图像需要强大的 GPU 来提供大量的计算力支持。但 GPU 长时间高负荷运行会严重影响生成图像的效率，容易引起功耗、续航甚至是发热问题。为了减轻 GPU 的工作负载，基于眼动追踪（Eye Tracking）的注视点渲染（Foveated Rendering）技术将会是未来 VR 设备中不可或缺的。

当 XR 设备捕捉到用户注视的焦点区域后，注视点渲染技术会把大量的工作负载集中到这个区域，只在这个区域以高分辨率渲染，并利用人类视觉较低的周边敏锐度在焦点周边呈现低像素密度，以低分辨率渲染。用户不会注意到外围的分辨率较低，因为人眼本身只能在中央凹处看到高分辨率。

如果 GPU 可以以全分辨率渲染只约 3 度左右的视野大小，那么提供的整体性能优势可以高达 20 倍左右。同时实现这一目标也需要零延迟的眼动追踪、极高的显示刷新率和高质量的重建算法，才不会发生画面的闪烁。Meta 头显同时支持固定注视点渲染（FFR）以及眼动注视点渲染（ETFR），通过应用程序性能测试，在默认分辨率下 FFR 可节省 26%-36% 的性能，而 ETFR 可节省 33%-45%。当默认分辨率降低 1.5 倍或更大时，所节省的 GPU 资源将更大。所以 ETFR 是降低带宽、

成本并提高画面质量的不二选择，当然眼动追踪以及注视点渲染技术也会持续的迭代以产生更优的方案。

3.3.5 透视技术

从当前 XR 设备发展看，目标是提供完美的虚实结合的交互体验。为了给 AR 和 VR 用户带来真正的 MR 体验，除了呈现在显示器上三维虚拟对象之外，还要具备叠加显示真实世界景象的能力，这个能力我们称之为透视技术。AR 和 VR 因为光学显示原理的差异，佩戴之后感知外界真实环境的方式有所差别。基于此，透视技术主要分为两种：光学透视（Optical See-through，简称 OST）和视频透视（Video See-through，简称 VST）。

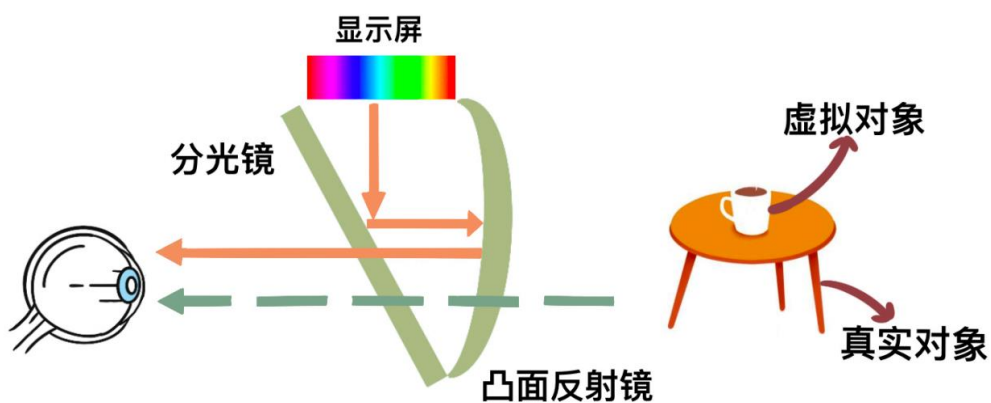


图 3.8 Optical See-through 透视技术

对于 AR 设备来说，通过光学镜片可以非常自然地看到外界真实环境的实景内容（如上图 3.8 中真实对象的桌子），AR 设备通过追踪环境中真实对象的位姿信息，渲染出一个三维虚拟对象（如上图 3.8

中虚拟对象的杯子），AR 用户通过光学系统可以观察到虚拟对象和现实对象在现实的三维空间中融合在一起，真切感受到了虚拟内容和现实世界的交互（如虚拟的杯子“真”放在现实的桌子之上），这种交互体验就是真正的 MR，让虚拟内容和周围的环境“融合交互，真假难辨”，所应用到虚实叠加光学效果，就是 AR 的 OST 光学透视技术。

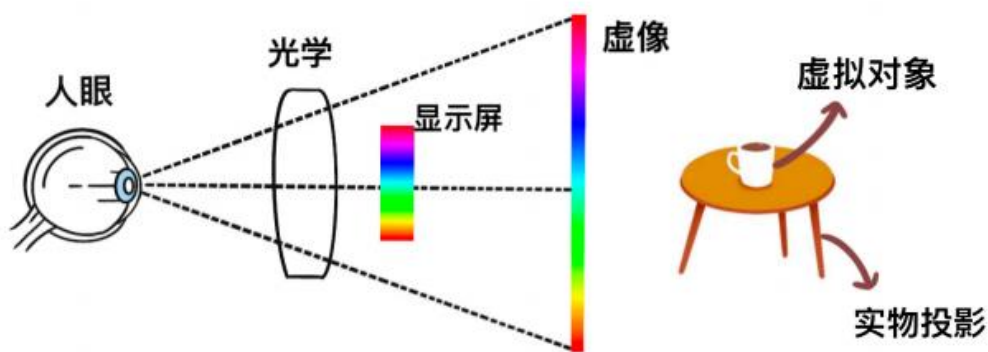


图 3.9 Video See-through 透视技术

对于 VR 设备来说，无法通过光学镜片看到外界真实环境的实景内容，只能通过摄像头拍摄或者重建一个跟实景相似的实物投影（如图 3.9 中实物投影的桌子），VR 设备同样通过追踪环境中的真实对象（如桌子）的位置信息，渲染出一个三维虚拟对象（如图 3.9 中虚拟对象中的杯子），用户佩戴后会发现渲染出来的虚拟对象和实物投影在虚拟空间中“融合”在一起，间接感受到了虚拟内容和现实世界的交互（如虚拟的杯子放在真实世界桌子的实物投影虚像之上），这是区别于 AR 的另一种 MR 体验，让虚拟的内容和周围环境投影产生了“融合”

交互”，这种虚“实”（环境的投影虚像）叠加的光学效果，就是VR的VST视频透视技术。

AR设备天然具备OST功能，而只有中高端VR设备实现了VST功能，使用一个单目摄像头可以大致实现外界场景的较为模糊的画面，MR体验不好；使用两个全彩摄像头的方案最为直接，但成本、功耗和算力的问题较难克服。配备了ToF传感器的VR头显可以更快速地捕捉环境的深度信息，便于重建更接近真实的透视体验。延迟是目前制约VR设备VST体验的一个关键因素，处理器必须对高分辨率的彩色相机捕获的图像进行编码，甚至三维重建，因此处理器的速度、带宽和功耗将影响整个VR系统的VST体验，对SoC要求相较手机会更高。

3.3.6 手势追踪技术

XR技术在不断发展的过程中，势必受到各种交互方式的影响。其中最常见也是最直接的交互方式之一就是手势交互。手势追踪技术在XR领域中扮演着非常重要的角色。

近些年来，随着计算机视觉与机器学习技术的发展，手势追踪技术又得到了新的突破。使用摄像头获取图像和视频数据，通过计算机视觉技术分析和识别手部的姿态，就可以实时获取手部运动的轨迹，从而实现高精度的手势追踪。目前，流程基本上分为四个部分：图像获取、手势检测、手势跟踪，以及手势姿态估计。

在算法方面，常用的手势识别算法主要分为基于模型匹配的方法和基于深度学习的方法。基于模型匹配的方法早期比较常见且广泛使用，该方法需要先建立手部姿态模型，再通过指定的特征点或者轮廓线匹配实际运动，但是该方法有明显的局限性，如难以应对部分遮挡和姿态快速转换的情况。现如今，基于深度学习的方法得到越来越广泛的应用，因为基于深度学习的方法能够获取足够多的手势数据，训练模型正确识别各种类型的手势，在实际应用中精度和稳定性比较高。

在芯片领域，手势算法可以在较强算力的 CPU 上执行，但有功耗大和帧率低的劣势。目前在 XR 设备上，一般通过高通的神经网络处理引擎（SNPE）或者数字信号处理器（DSP）硬件加速模块来提升运算速度、降低负载和功耗。值得一提的是，Hololens 2 专门设计的 HPU（Holographic Processing Unit）芯片包含了多个硬件加速计算单元，可以对手势进行分析和识别的同时，大幅降低延迟。

总体而言，随着计算机视觉和机器学习技术的逐步成熟，以及拍摄设备的不断升级，现代手势追踪技术已经取得了许多突破，并且手势追踪技术也逐渐得到了更加广泛的应用。在 XR 领域中，手势追踪技术已经相当成熟，可以应用于虚拟手势交互、虚拟操控、虚拟人物表演等多个领域。当然，手势追踪技术在交互流畅性、响应速度等方面还有一定的提升空间，未来随着技术的不断进步，相信手势追踪技术也会更加完善和普及。手势追踪技术的发展也带动了芯片行业的创

新和进步，催生了更多高效、低功耗、多功能的芯片产品的诞生。因此，手势追踪技术和芯片行业是互相促进的一种关系，随着技术的不断发展，这种关系也会更加密切。

3.3.7 手柄追踪技术

扩展现实技术正在迅速发展，手柄作为一种交互配件随之而来，虽然许多 VR 设备支持了比较好的 6DoF 手势交互，但手柄的交互体验目前更加准确和高效。为了实现高品质的 VR 游戏和体验，手柄追踪技术成为了必不可少的一环。

手柄追踪技术基本原理是通过传感器获取手柄的运动轨迹，然后将数据传递到计算机进行图像渲染，以模拟出精确的运动感受。目前主流的手柄追踪技术主要是基于视觉的追踪技术。基于视觉的 6DoF 追踪技术可以分为两类：手柄自主追踪技术和手柄被动追踪技术。目前主流的 VR 设备使用的是手柄被动追踪技术，这种技术主要通过手柄上特殊标记信息被 VR 眼镜上的摄像头追踪，实现 6DoF 功能。例如，HTC Vive 手柄采用了被动反射式红外光标记和 Lighthouse 红外激光定位系统，通过定位和分析反射点的位置，计算出手柄的位置和方向；Oculus Touch 采用了基于红外 LED 的跟踪技术，通过外置红外摄像头识别手柄的运动。但像 Quest Pro 的这类设备，另辟蹊径，采用手柄自主追踪技术，在手柄上增加了主控芯片和 3 个摄像头用来做自主 6DoF 定位追踪。

手柄是一种能够追踪用户手部位置和动作的控制器，对 XR 交互有很大的价值，可以从以下几方面带来体验的提升：

1) 更真实的交互体验：6DoF 手柄可以追踪用户的手部位置和动作，为用户带来更真实、更自然的交互体验。用户可以自由地在虚拟环境中移动、抓取和放置物体。

2) 更高的精度：6DoF 手柄的追踪精度比传统的控制器更高，可以捕捉到更细微的手势和动作。这为用户提供了更多的控制力，使得交互更加准确、自由。

3) 多样的交互方式：6DoF 手柄追踪精度较高，因此在交互方式上，可以更加丰富多样。例如，用户可以使用手指点击、捏持、抛出和抓住物体，或者画出形状来进行交互。

被动追踪手柄对芯片的算力要求并不高，一些低功耗低算力的芯片，例如 NXP IMX6 系列处理器、STM 的 STM32 系列处理器、Nordic 的 nRF52 系列芯片就可以满足需求。但是基于主动追踪技术的 Quest Pro 手柄，却至少需要高通 6 系列的中端芯片才能提供足够算力支持，随着 VR 技术的不断发展，手柄芯片也在不断地迭代升级。

第四章 扩展现实设备底层芯片需求

随着新一代移动通信规模商用，高度沉浸、多感官同感、人机融生的呈现需求显现，扩展现实设备相比传统智能终端具有更复杂的交互功能，空间计算和云计算、人工智能、三维建模等更强的软硬件能力。因此，需要通过高集成度高性能的芯片设计、高效的芯片设计验证和成熟先进的制造工艺实现对扩展现实头显的核心计算芯片、手部配件核心计算芯片、其他芯片的创新研发，来满足扩展现实多样化的呈现需求。

4.1 核心计算芯片能力

扩展现实头显设备需拥有多种类型芯片组合，主要包括：核心计算芯片、连接芯片、光学驱动芯片、FPGA 可编程芯片、射频芯片、音视频编解码芯片、音频功放芯片、电源管理芯片、存储芯片以及其他辅助芯片等。XR 头显设备底层芯片构成如图 4.1。

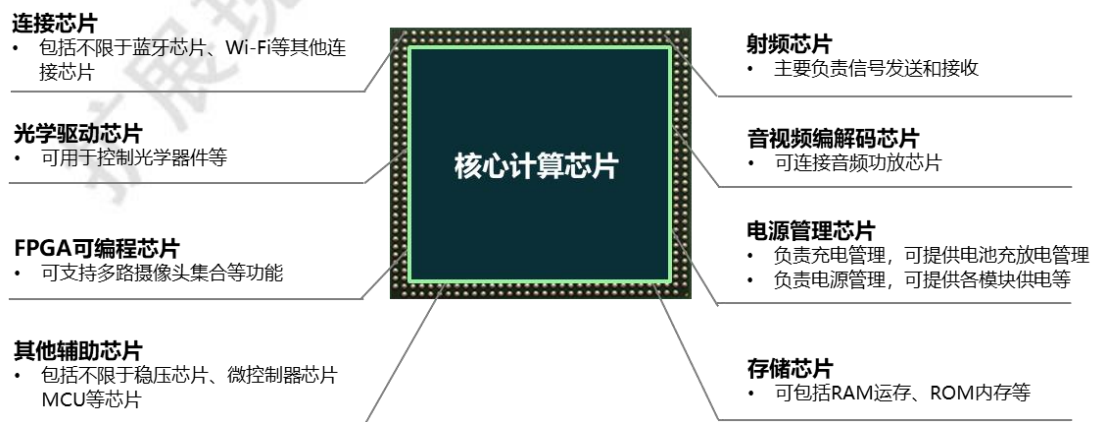


图 4.1 XR 设备底层芯片构成及功能

根据扩展现实头显设备特有功能和更高性能要求，其核心计算芯片正向 XR 专用芯片发展。核心计算芯片 SoC 是 XR 产品实现运行控制和数据处理的核心，早期 XR 产品多采用移动消费级芯片，但 XR 设备对芯片有更多特有需求，例如更高算力以支撑高品质图像处理、丰富交互功能、多传感器信息融合，功耗和散热。

核心计算芯片需要满足 XR 设备在感知、计算、功耗、性能、连接、AI 等方面的要求，并提供操作系统高效运行、多传感器信息融合处理等支撑，其主要功能模块包括：CPU、GPU、ISP、VPU、WCN、NPU 等。

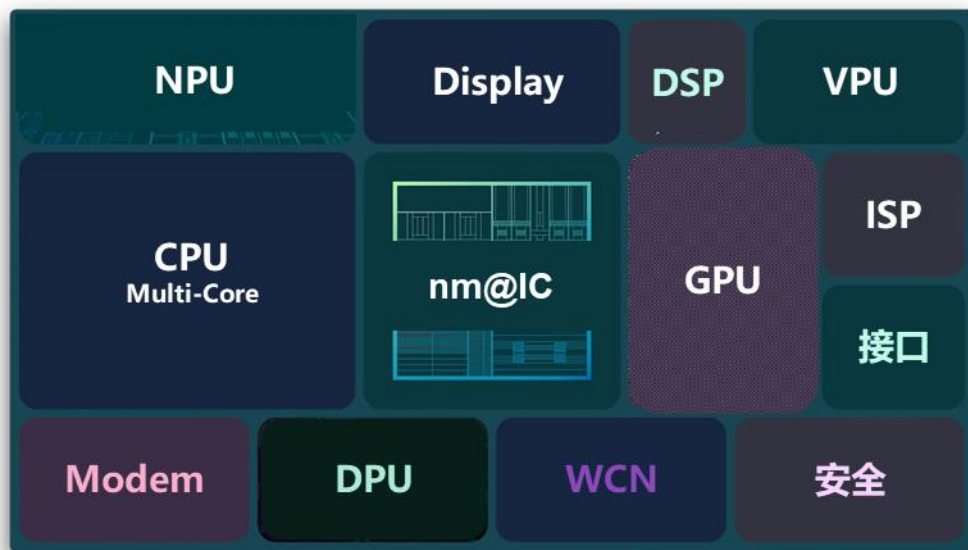


图 4.2 核心计算芯片组成

4.1.1 计算能力

XR 设备芯片比手机芯片需要更强的空间计算能力、计算渲染能力、高效算力协同。计算单元架构融合可优化数据搬运、降低功耗、提升效能。多域计算以及中央计算的技术演进迭代，可以完美满足 XR 芯片计算架构的技术演进。

CPU 能力

中央处理器（Central Processing Unit, CPU）是核心计算芯片的核心部件之一，主要用来解释计算机指令以及处理计算机软件中的数据。

对 XR 领域增长的计算需求，对 CPU 的计算架构提出了很大的挑战。既要兼顾性能来提升用户体验，也要有强大的功耗控制能力，应对未来 XR 小型化以及轻量化的趋势。XR 核心计算芯片 CPU 至少满足四核 A7X 及以上，且主频 2GHz 以上能力，且需要更先进的生产工艺。

未来更高级别 CPU 架构可提供更加强大的 AI 能力和 DSP 能力，为 XR 应用算法实现提供良好支撑并为运算实时负载均衡提供可能。另外，为了 XR 内容生态开发提供便利，CPU 需提供强大的生态和软件支持。

GPU 能力

图形处理器（Graphic Processing Unit, GPU）是一种特殊类型的处理器，将原本由 CPU 完成的图形绘制运算等工作改用专门的图形图像处理器来完成，成百上千个内核经过优化以后做并行运算，可以大幅提升运算性能，并节省 CPU 资源。GPU 不光可以用于图形图像渲染，还可用于运行分析、深度学习和机器学习算法。

正是由于 GPU 强大的计算机图形处理能力和并行计算能力，才能实现 XR 渲染复杂的内容运算。为保证 XR 内容流畅显示和高清屏驱动，以及 XR 基本像素填充率、纹理速率等，需要 GPU 高速计算支撑，GPU 能力至少支持 512 GFLOPS 图形渲染能力，支持 singlebuffer, multiview2, FOV 渲染，context 小粒度抢占等。

专用 VR 芯片需要对 CPU、GPU 进行功耗优化，采用更小制程实现算力提升，GPU 先进性微架构将向着更多更精细的图形渲染能力、更多更高效的通用计算能力，更智能的 AI 运算能力方向迈进，以提供更强的图形处理和视频渲染能力，并提供光线追踪等算法能力。大规模扩展计算能力的高性能计算、人工智能计算、更加逼真的图形展现方向将成为 GPU 未来发展的三个趋势。

NPU 能力

NPU 是神经网络处理单元--Neural Processor Unit 的简称，本质是一个 AI 加速器，专为 AI 加速设计的定制电路，并且包含一些必

要的控制单元和算法，来执行机器学习算法，解决传统芯片在神经网络运算时效率低下的问题。

传统 CPU 和 GPU 也可以进行神经网络相关的运算，但是神经网络运算涉及到较多的乘加运算，而且是有数据相关性的乘加运算。传统的 CPU 架构进行类似的累加计算，效率非常低。GPU 也可以完成神经网络运算任务。但针对神经网络运算特殊优化过的 NPU，性能上比 CPU/GPU 要高得多。

目前主流平台均达到 2.4TOPs 以上 AI 算力。为满足高效高品质的图形渲染，AI 算力或将需要达到 15TOPS 乃至更高水平。未来 NPU 对感知能力的提升，可能会带来人机交互方式的改变，具备更高灵活性、实用性的通用智能芯片发展是其技术发展的必然方向。

VPU 能力

VPU(Video Processing Unit, 视频处理单元) 是用于压缩和解压缩数字视频的专用电路。每秒 30 帧，每像素 24 bit，分辨率是 1920*1080 的视频，无压缩时需要带宽为 1423.82 Mbps。在存储空间和联网速度有限的情况下，要实现高清视频的传输就需要用到数字视频压缩和解压缩技术。

VPU 能力方面，需要支持编解码 HEVC, VP9, AVS2, H.264, 最大延迟小于 16ms。芯片编解码器支持多 slice/tile, 支持 slice/tile 低延迟输出。编码器支持灵活稳定的码率控制，使码率能够平滑输出。

多核解码器能支持多核帧内并行解码，以降低解码延迟。为消除 VR 设备纱窗效应，芯片至少具备 4K@60fps 解码基本能力。由于 AR 技术实现不同，清晰度和帧率要求，较 VR 低一些。

理论上来说，人类视网膜中央凹最大能达到 60PPD 的可视度，在水平 120 度与垂直 135 度的 FOV 下，双眼需要约 1.2 亿像素数，未来芯片理想状态是满足 16K 渲染能力，至少 72Hz 的可变刷新率。

DPU 能力

DPU 是显示处理单元，其功能比较固定，一个是 2D 加速（包括缩放、合成等功能的支持），一个是显示输出。

对于扩展现实设备来说，随着输入分辨率的提高对 DPU 的能力要求也越来越高，要求的运行频率也更高。现有扩展现实芯片能力需能驱动双屏，单屏最好满足 2.5K@120Hz 或者 4K@90Hz，需要支持高动态光照渲染（High Dynamic Range, HDR）和可变刷新率（Variable Refresh Rate, VRR）。

随着分辨率的提升，带宽要求也更大。单眼 4K 分辨率意味着 DPU 要支持 8K 的输入，带宽压力很大，目前的通常做法是切分成两个 2k，这就是 split 功能。未来要支持单眼 8K，DPU 需要支持 16K 的输入，可能需要不止一个 DPU 的支持。带宽的高要求也对视频压缩能力有需求，减少内存带宽占用，特别是与 GPU 的交互带宽。

DSP 能力

数据信号处理 (Data Signal Process, DSP)，指对信号（例如音频信号）进行数字化处理以达到目的。在 XR 中，DSP 可以用于处理图像、视频和音频信号，以提高质量、性能和效率。例如，DSP 可以用于实现空间音频、视频编解码、机器视觉算法等。

XR 中的 DSP 性能要求取决于所需的计算量和实时性。例如 AR 的实现涉及一系列计算，探测真实物体、计算物体的空间位置和方向、计算虚拟物体叠加的位置、渲染虚拟物体等，为避免眩晕实时显示，其对计算过程时间有较高要求（一般不超过 20ms）。因此，XR 中的 DSP 性能要求较高，需要具备较强的计算能力和低延迟。未来面向 XR 中高速、高密度数据处理的应用，DSP 向更低功耗和更高效的方向迈进。

ISP 能力

图像信号处理 (Image Signal Process, ISP)，在相机成像的整个环节中，它负责接收感光元件 (Sensor) 的原始信号数据，主要对前端图像传感器的输出的图像信息做后期处理和计算，主要功能有色彩插值、自动白平衡、颜色校正、色彩空间转换、自动曝光控制等，XR 依赖于 ISP 才能在不同的光学条件下都能较好的还原现场细节。

由于 XR 需要三维重建、视觉定位、插帧等多重技术的支撑，现有 XR 主控芯片需支持多路 camera 同时曝光，camera 总输入能力提

升成为趋势。未来 ISP 随着高像素采样和多摄的不断发展，将会提出更高的性能收益。

4.1.2 无线通信能力

5G、Wi-Fi 是主流传输技术，无线传输的延迟是个最大门槛。XR 设备需要分布式计算处理、实时互动等特性支撑，且无线传输存在编码、传输、解码等环节，所有通路处理和传输时延会大幅增加，对无线传输的带宽和时延要求提高。

Modem

Modem 提供调制解调能力，联合射频芯片完成蜂窝无线信号的处理、接收和发射。5G 可提供更低时延、100MHz 带宽乃至更高能力，为 XR 的数据传输提供良好通道。

WCN 和其他通信能力

XR 芯片可包括 WCN 模块，具备 Wi-Fi 等能力，满足 XR 业务高频多连接并发，聚焦云网边端的协同渲染，也可辅助其他短距通信能力。

4.1.3 芯片接口

XR 设备主控芯片与计算单元间的无线数据传输，对传输速率、接口有着较高要求。未来 XR 芯片需要满足高速及大容量的存储需求，支持更大的总线频率带宽，支持高速扩展接口等。

4.2 手部配件核心计算芯片

扩展现实手柄设备主要由存储芯片、核心计算芯片、电源管理芯片、IMU 芯片、蓝牙芯片、射频芯片、Wi-Fi 芯片、马达驱动芯片等组成。

在一些高级 XR 设备中，手柄有专门的核心计算芯片。手柄核心计算芯片能力较头显核心计算芯片低很多，需要对传感数据的简单处理和反馈。根据手柄作用的不同、形态的差异，需要一定程度的传输、定位、传感能力，并向自带摄像头的独立手柄方向迭代。

4.3 其他周边芯片

除了核心芯片和手柄核心计算芯片之外，XR 用到的其他芯片基本都可以与其他场景下的芯片复用，包括不限于连接芯片、光学驱动芯片、FPGA 可编程芯片等。连接芯片包括蓝牙芯片、Wi-Fi 芯片等；光学驱动芯片用于光学显示和双目光学控制；FPGA 可编程芯片可用于负责多路数据集合到一个输入接口；射频芯片用于将 Wi-Fi、蜂窝通信发射功率放大；音视频编解码芯片用于支持复杂的音视频信号处理应用和低功耗应用，音频功放芯片保护扬声器工作正常；电源管理芯片主要负责给外围 IC 供电、CPU 供电、电池充放电管理、电量管理，让电源更平稳；存储芯片包括 ROM 闪存芯片和 RAM 内存芯片；其他芯片包括稳压芯片、微控制器芯片、快充协议芯片、升压芯片、电压电平芯片等。

4.4 芯片其他需求和芯片制造

4.4.1 交互感知

XR 设备芯片的多维感知能力，从视觉感知、听觉感知进步到触觉、嗅觉、味觉、位置、距离、速度、角度、动作、环境等方面的感知，多维的感知能力帮助 XR 完成物理世界的数据采集，更好的提升用户认知沉浸感。硬件方面，更丰富的感知交互功能要求更多传感器和更强算力芯片的参与。

交互方面，除支持多路并发摄像头，未来需支持高精度实时追踪用户的头部、嘴唇和眼球，并支持手部骨骼追踪等其他追踪算法的开发，搭载更多硬件加速模块支撑更复杂算法，因此需具备眼动追踪、表情追踪、语音输入、3D 建模、注视点渲染等众多新型复杂交互功能。

4.4.2 芯片功耗

XR 设备的续航能力影响着舒适体验，在保持芯片高算力的同时，需要兼顾功耗和散热，以实现较好的续航能力。低功耗、高能效比、更长的续航时间成为未来 XR 设备的技术趋势。

4.4.3 芯片安全

活体检测、人脸解锁等，未来 XR 虚拟资产以及用户隐私的保护需要芯片提供强有力的安全支撑。芯片具备安全可信运行环境，即能提

供安全性和完整性保护的独立处理环境，如划分为安全区域和非安全区域，满足内容和信息保护需求。

4.4.4 芯片制程

摩尔定律依然适用于当前集成电路发展趋势，高芯片制程有助于实现高集成、高速运算和低功耗。为满足 XR 高数据计算量、数据处理、传输需求，芯片采用先进制程工艺势在必行，成熟的工艺制程也可以保证芯片的出货量和良品率。目前主流 XR 芯片可达到 7nm 乃至更高，未来 1~2 年可能会跳跃式发展至 4nm。

4.4.5 芯片封装

在先进制程芯片的封装中，多采用多维异构封装技术，并已发展为晶元级、系统级、Chiplet、3D 封装技术，需克服封测过程的连接和散热等挑战，实现更高密度的芯片封装。

第五章 面向扩展现实的 IP 和芯片平台

5.1 现有 IP 如何满足 XR 芯片需求

5.1.1 CPU

目前移动端对 CPU 的需求越来越高，Arm CPU 的性能也在每年不断提升，除了在架构上的升级外，每年 CPU 的性能也在不断增长，以应对不断增长的计算和性能需求。尤其对扩展现实领域，对 CPU 的计算架构和微架构都提出了很大的挑战，既要兼顾性能来提升用户体验，也要有强大的功耗控制能力，以应对未来扩展现实小型化以及轻量化的趋势。

Arm CPU 的架构这几年从 V8 全面过渡到 V9 的时代，除了每代增加性能外，也提供了先进的 ML 运算能力以及 DSP 能力，来为未来扩展现实的 workload 的支持提供可能性。

从最早 V7、V8 时代的 NEON™ SIMD 的引擎到现在 V9 的 SVE2，Arm 对未来的 SIMD 的 workload 的理解有了一个重新定义。有了 SVE2 的加持，用户或者软件开发人员可以更多的依赖于编译器去做自动的向量化而不再需要通过传统 NEON 手写汇编的形式，进一步加快软件开发者开发程序的效率；同时，SVE2 在不同平台上的部署可以做到无缝的迁移，开发的基于 SVE2 的程序，都可以做到无缝的迁移，不需要重新编写，也为未来扩展现实应用的云端开发部署提供了极大的便利，极大地提升了未来扩展现实应用的开发部署效率。更进

一步的, SVE2 还额外增加了很多 DSP 指令的支持, 以及更精细的 SIMD 计算的控制编程模型, 为未来的扩展现实的应用算法的实现以及部署都提供了良好的支撑。

同时, Arm 每代 CPU ML 运算能力的提升也源自于 Arm 架构的持续升级, 从 Arm® v8.4-A Dot product, Arm® v8.6-A Int8 & BF16 Matmul, Arm® v8.6-A Bfloat16 到 Arm® v9-A SVE2 的这些架构特性的引入, 都对 ML 的性能提升有很大的助益, 相信未来对扩展现实算法, 特别是对一些感知, 融合交互算法的性能提升都有很大的帮助。

用户的隐私以及安全保护一直是 Arm CPU 重点工作的方向, 在 Arm V9 架构上引入的新的安全隐私保护机制, 无疑对未来虚实现实资产以及用户隐私的保护提供了一个强有力的支撑。

Arm CPU 除了 CPU 核心本身之外, 其基于 DSU 所实现的 DynamIQ™ 大小核方案, 也对未来扩展现实多样的性能需求以及功耗控制提供了灵活和强大的硬件支撑。

最后, Arm CPU 强大的生态以及软件支持, 可以快速的帮助用户去开发基于扩展现实的应用, 去实现自己的想法。

在一个典型的扩展现实芯片里, 一般采用大核+小核的方式, 目前 Arm 可用于移动终端设备的 V8 架构最新大核是 Cortex®-A78, 小核是 Cortex®-A55, V9 架构最新超大核是 Cortex®-X3, 大核是 Cortex®-A715, 小核是 Cortex®-A510。

5.1.2 GPU

对于扩展现实来说，对 GPU 的需求要求很高。我们可以看到目前的扩展现实的用户体验都或多或少和 GPU 的性能息息相关。对于扩展现实这个市场来说，GPU 往往要应对高分辨率以及高刷新率的渲染，除了 GPU 性能要足够强大以外，对其功耗控制以及散热的管理也同样重要，这些也会对用户的最终体验以及长时间佩戴有很大影响。

为了解决扩展现实的痛点，Arm 的 Mali™ 系列以及 Immortalis™ 系列 GPU 在硬件以及软件上加了很多新的功能来支持这个市场。首先，Arm 的 Mali GPU 支持 Multiview 的 extension，可以同时去渲染左眼和右眼的图像。并且，Arm 的 Mali GPU 在注视点渲染上也做了增强，在降低 GPU 的 workload 的同时去提升系统的能效，并且可以有效地和眼动追踪相结合，来进一步提升 GPU 的能效。

对于扩展现实而言，MSAA (Multisampling Anti-Aliasing, 多重采样抗锯齿) 尤其重要，因为用户需要近距离的观看屏幕，渲染上细小的瑕疵都能被用户捕捉到，Arm GPU 原生强大的 MSAA 算法，可以在保证性能的同时提升渲染图像的质量，最终提升用户观感。

5.1.3 NPU

在扩展现实芯片中，NPU 可用来实现算法加速，包括物体检测、识别、眼球追踪、插帧、超分等。通过算法与硬件的结合，可以将一部分原本由软件运行的任务交给 NPU 硬件加速，大大提高运算效率。

Arm China 的“周易”NPU IP 是针对深度学习的 AI 专用处理器，采用可编程“周易”架构，支持从 1 TOPS 到上百 TOPS 的算力配置，提供更高的精度和灵活性，可支持多种数据精度融合计算。“周易”NPU 可针对扩展现实场景做优化，可满足扩展现实运算中对插帧等需求的支持，同时采用 i-Tiling 技术大幅减少带宽需求；支持 int4/int8/int12/int16/int32，fp16/bf16/fp32 等多种数据格式，可实现多精度融合计算；支持 TSM 任务分拆和管理技术，充分发挥各个计算单元效能；支持自定义算子，满足各种模型部署需求；支持定制差异化 AI 解决方案。同时“周易”NPU 提供一套完整的人工智能软件方案，使开发者可以方便、快速地进行算法移植和调试；支持主流的人工智能框架，如 TensorFlow、Caffe、ONNX、PyTorch 等模型；提供多种开放接口，支持用户模型和自定义算子等开发和调试；支持 Android、Linux、RTOS、QNX 等不同 OS；支持 TVM、Arm NN 等 SoC 异构计算，从而有效使用 CPU、GPU、NPU 等计算资源；Bit 精度的软件仿真平台，便于算法移植和部署。

5.1.4 VPU

视频编码技术自诞生以来，不断更新换代，致力于更优的编码效率和压缩效果，而更高的压缩效率来源于更高效的编码算法。假设使用 35:1 H.264 编码来存储，每秒所需要的速率为 $1423.82\text{Mbit}/35 = 40.68\text{Mbit/s}$ ，也就是 40M 左右带宽即可满足传输要求。

Arm China 的“玲珑”视频处理器是一款为满足主流市场不断增长的 4K/8K 实时编解码需求而设计的高效视频处理器。这款全新视频处理器针对主流市场的视频流媒体技术进行了大量投入，通过一系列智能权衡实现了极大优化，为所有合作伙伴提供灵活的组合和选择。“玲珑”V6/V8 视频处理器的设计旨在确保整体系统的成本、功耗和面积最优化，能够显著节省存储空间，同时在更小带宽内实现最高品质。“玲珑”视频处理器系列还提供多实例、全方位的软件及固件工具。

5.1.5 DPU

有鉴于 VR 带动市场对高级设备屏幕的需求逐渐上升，4K 以上分辨率以及更高的帧率，衍生出不少系统效能方面的难题。对此，采用专用硬件组件 DPU 来执行合成、旋转、高质量缩放以及其他图像处理，可有效降低 GPU 的工作量。和 GPU 必须进行多次读取、处理与输出，造成整体功耗较高相比，DPU 可以处理完多项任务后进行压缩与一次性输出，有效降低系统功耗。ASW/ATW 相关功能如果放在 DPU 上做，可以节省带宽，降低延迟，也是未来的一个发展方向。

Arm China 的 D71 显示处理器可支持 4K120FPS 的画面输出。Mali-D71 和 Hardent 公司的 VESA DSC 组合解决方案，可以在既定的移动功耗范围内，通过视觉无损压缩将传输带宽减少 3 倍，从而实现更加身临其境的 VR 体验。

5.1.6 ISP

在扩展现实芯片中,ISP 要负责对接入的多个摄像头数据的处理,4 个 SLAM 算法摄像头,1 个 See Through 摄像头,2 个眼球追踪红外摄像头,这就要求 ISP 至少支持 7 路摄像头数据输入。随着对算法精度要求的提高,需要的摄像头数量和分辨率都会进一步增加。

Arm China 的“玲珑” i5 ISP 具有高画质、低延时、高兼容性、易扩展、面积精简和低系统带宽等特点。硬件架构灵活可配置,客户可自行选配可选模块进行集成;其多元的工作模式可兼容线性、原始/压缩的 HDR 数据,支持单路及多路摄像头输入,支持超高分辨率分屏处理;DMA 接口数据输入输出模式可配,可在 ISP 多个节点输出不同格式的数据;软件 API 接口丰富,图像效果调试流程简易清晰;提供了丰富的软硬件参考设计,如标定工具、调优工具和 MIPI 转 DVP 数字电路等。

5.1.7 SPU

SPU 是芯片里有关安全相关特性支持的 IP,包括信任根的管理,加解密相关 IP,数据隔离通路,安全启动,安全调试等相关功能的支持,目的是保证系统从启动开始就是安全可信的,构造一个可信的执行环境,并避免数据在运行过程中被其他恶意程序窃取。

在扩展现实芯片中,需要实现用户身份认证,支付,数字版权管理等,这些都需要安全相关 IP 的支持。

Arm China 的 “山海” S12 提供一站式安全解决方案，包含硬件加解密引擎，软件和安全服务，从芯片的安全 IP 层到云端安全应用和安全管理进行全链路的安全保护。“山海” S12 的核心模块 TrustEngine-600 是密码算法引擎，其包含国际通用算法以及中国商用密码算法。“山海” S12 硬件算法引擎默认支持 TrustZone®，并且其算法及安全能力可以根据需要灵活配置。此外，客户既可以使用安谋科技提供的随机数方案也可以自定义随机数方案。“山海” S12 和其他 IP 共同打造的可信执行环境能够为需要安全保护的应用和逻辑提供一个安全的运行环境，确保敏感数据和安全业务在隔离、可信的环境中存储、执行和保护。“山海” S12 为 AIoT 系统提供从 SoC 硬件 IP 到云端安全服务的完整的安全解决方案，可以为多种安全解决方案如数字版权保护、AI 安全、身份认证等提供基础安全能力。

“山海” S12 还提供云端的设备管理以及安全升级应用，为 OEM 和云服务商快捷实现设备管理和固件升级提供安全部件。“山海” S12 丰富的固件支持为合作伙伴的快速集成提供了保证，帮助合作伙伴快速设计芯片并快速推向市场。“山海” S12 采用模块化设计，可灵活配置，合作伙伴可以根据自身安全需求组合出性价比最高的安全解决方案。

5.2 芯片原型平台

5.2.1 一体式 XR 芯片原型平台

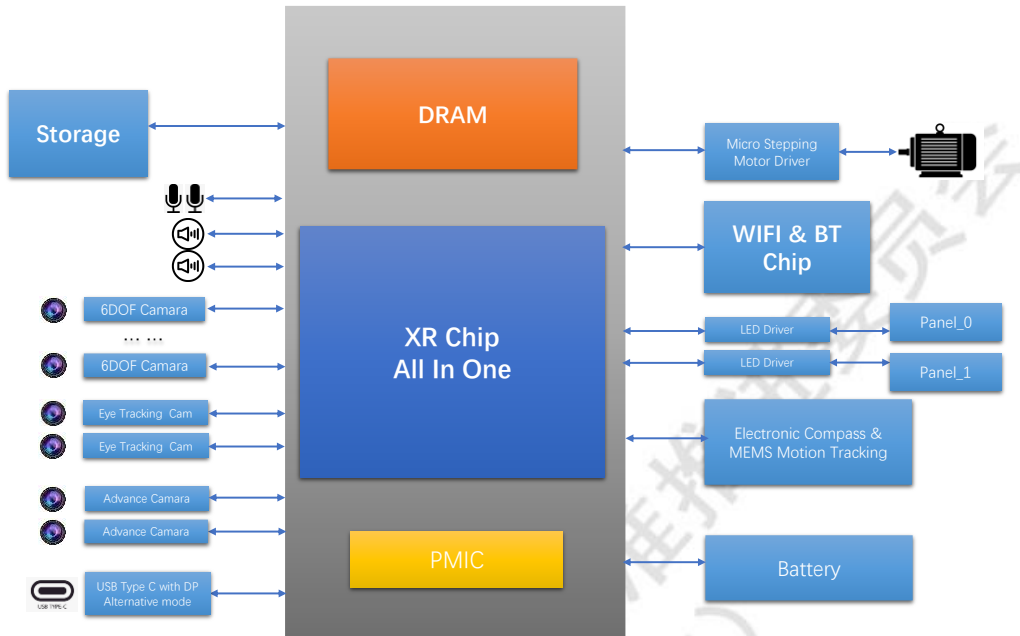


图 5.1 一体式 XR 芯片原型平台

随着扩展现实领域的技术进步与产品日趋成熟，今后对于扩展现实专用芯片的需求也日益强烈。同时为得到更好的沉浸式的扩展现实体验，大视野，高分辨率，高刷新率，高性能图形渲染，空间定位与感知，低数据传输延迟，低网络延迟，低显示延迟，无一离不开扩展现实芯片的进一步发展。因此相对于传统手机类移动芯片，扩展现实专业芯片的需求无疑有着更为苛刻的性能，功耗，面积要求。

对于一体式芯片 (VR)，单芯片 (Single Die) 平台如上所示，涵盖以下系统功能：

扩展现实芯片平台中，多路 6DoF Camera 通过 MIPI 与 XR 主芯片相连，传感器和感应器芯片通过低速外设与主芯片中的 Sensor Hub 相连，为 SLAM 计算提供主要输入。而作为透视功能重要输入源的高级视频透视摄像头则通过 MIPI 为主芯片提供现实世界高分辨率视觉图像。眼球追踪摄像头（Advanced Eye Tracking Camera）也通过特定的 MIPI 通路为主芯片相连，为 GPU 渲染提供眼部追踪数据，为随后 GPU 的注视点渲染提供重要输入。平台中具有音频编解码（Audio Codec）、音频模数/数模转换（Audio ADC/DAC）和麦克风输入（Mic In）功能，提供完整的高集成度的多媒体服务。芯片内置的 Sensor Hub 单元提供对瞳距调节马达，线性马达的控制支持，以及相应的传感器服务供主 CPU 调用。扩展现实芯片平台中提供完善的高速外设接口，如支持 DP（Display Port alternative mode）的 USB Type-C 接口，提供 TX 方向影像输出与输入，2~4 lane PCIe Gen4 接口服务于外部 Wi-Fi 芯片控制与网络数据传输，亦或提供额外 PCIe 接口与外置的存储器、协处理器芯片进行互联扩展。芯片中具有高性能 DPU 来支持单眼 4K，双眼 4Kx2 级别的高帧率影像输出，且具有 ATW，ASW 帧补偿，可变帧率，HDR 等扩展现实必须的图像处理功能。内嵌的 PMU（Power management Unit）可直接控制外部 PMIC 芯片，完成高集成度，简化上下电与低功耗控制流程，芯片平台可通过控制 PMIC 芯片对电池进行电源管理。芯片平台兼容 Multi DDR protocol DRAM controller 如 LPDDR4/4x/5，提供多样化广泛的带宽支持能力。

芯片主体内具有高性能 CPU/GPU/VPU/ISP/NPU 等多媒体/AI/控制 IP 子系统，以及先进同步/异步总线，以及与系统匹配的 SMMU 机制，以及多样化、多手段的低功耗策略。专用的 SPU 子系统负责芯片内的安全启动，支持安全系统，完成安全与非安全世界数据控制分离，通信。整体芯片平台还需开源与硬件系统匹配的，且可被不同用户修改的底层软件栈，如系统启动固件(system boot firmware)，应用处理器固件（AP firmware），安全管理（Security management），boot flow，Kernel/OS。方便扩展现实芯片平台发展演进的一致性与生态环境搭建。

对于现阶段的扩展现实技术发展的多样性，与市场投入产出比的理性考量，多芯片组合共同完成扩展现实芯片平台成为目前普遍的共识。针对一体式芯片(VR)，Chiplet/Multi Chip 平台如上所示，Main SOC Die 和 Co-processor Die 通过高速 I/O 口进行互联，可支持 Multi Chip (PCIe) / Chiplet (UCIe)。

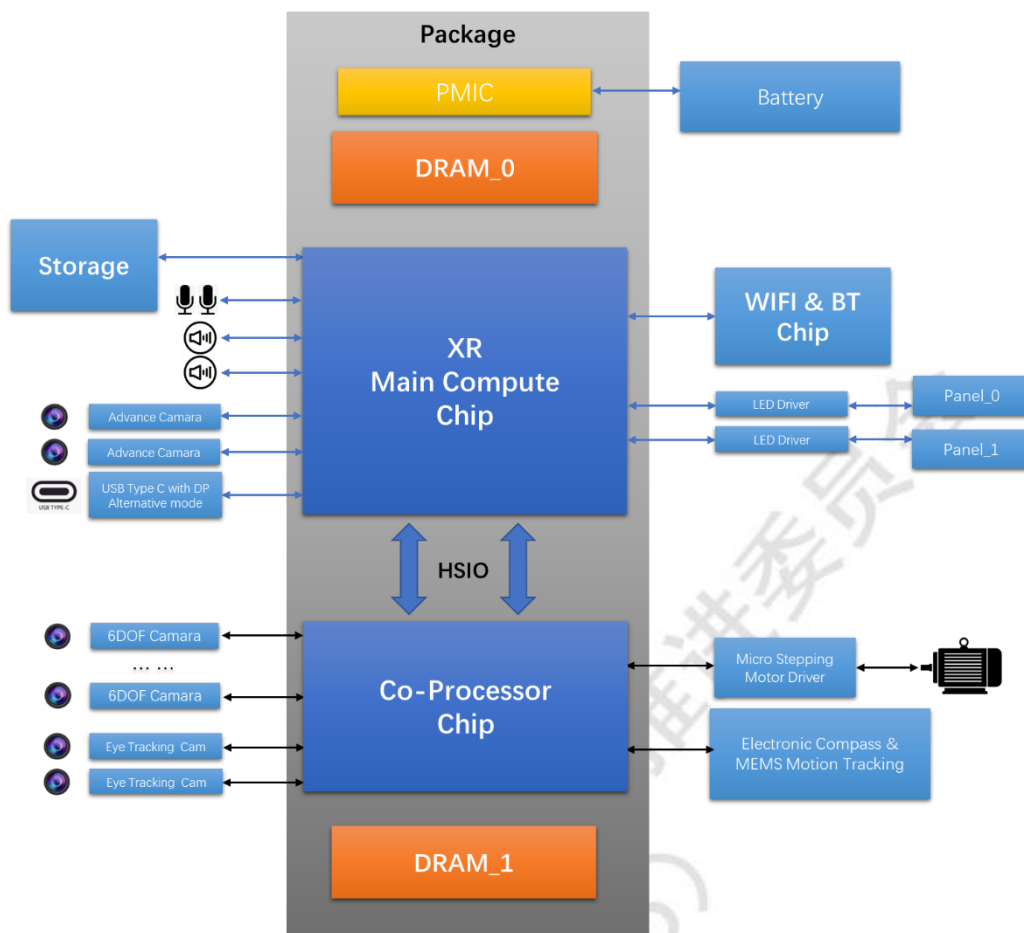


图 5.2 多芯片 XR 芯片原型平台

作为面向传统手机/PAD/PC 市场设计的主计算芯片 (Main Compute SOC) 负责整体系统控制，调度分配控制主计算芯片和协处理器 (Co-processor) 之间的任务 (Job)，XR APP 以及主要服务与 OS 运行于 Main SOC 中，其主要负责高密度计算，内置高性能 NPU/GPU/CPU/VPU/ISP/DPU 介入的任务，如 SLAM 后端算法/loop/optimization, Graphics Rendering 都由 Main compute SOC 负责。Wi-Fi 芯片与其直接互联，主 SOC 可通过 Wi-Fi 负责与芯片外部网关，Internet 设备通信。内部高性能 GPU/DPU 负责渲染以及显示合成后的双目高分辨率图像。主 SOC 负责接收高分辨率摄像头

(High Resolution Camera) 的影像数据流, 高像素图片捕捉, 通过 UFS/PCIe 高速接口主芯片可与外部存储芯片进行数据交互, 内置独立内存控制器(Memory Controller), 和专属内存访问通道, 可提供完整多媒体支持, 如编解码, 音频服务等, 为系统平台整体的低功耗策略与控制提供接口, 对外部 PMIC 进行完整控制, 提供主控 SoC 初始化配置和协处理器初始化通信引导服务, 芯片平台级的安全管理与协处理器安全通信。

多芯片组合完成扩展现实平台任务中的协处理器 SOC, 作为对传统手机/PAD/PC 市场设计的主 Compute SOC 芯片的补充, 负责协助主芯片实现完整扩展现实设备所需芯片功能。其主要作为设备角色受 Main Compute SOC 控制。协处理器具有单独初始化, 安全启动 (Security Boot) 能力以及必要安全功能。协处理器内部的 AI/DSP IP/子系统可负责轻量级算法, 轻量级物体识别, 人脸识别, 物体追踪能力。具有适量的 SLAM/VIO 前端算法实现能力, 来降低主 SOC 系统负载, 简化数据交互流程。对于负责接收的 6DoF 与眼球追踪摄像头传输过来的图像数据, 可进行预处理。内建 Sensor HUB 控制单元负责 IMU、红外、线性马达控制以及系统时间戳功能。协处理器具有多样性的低功耗电源管理与唤醒能力 (如语音识别, 手势识别, 图像识别唤醒系统)。独立内存控制器, 提供与协处理器角色匹配的带宽, 且与主控 SoC 内存栈隔离, 简化系统复杂度。亦或者通过 Chiplet 方式与主 SOC 共享 DRAM, 进一步降低系统面积, 加大系统集成度。协处理器具有独立 Power 控制权限和能力, 且可受主控 SoC 电源管理

单元管理，有功耗控制消息（Power message）通道唤醒进入低功耗（low power）模式的主控 SoC。

5.2.2 分体式 XR 芯片原型平台

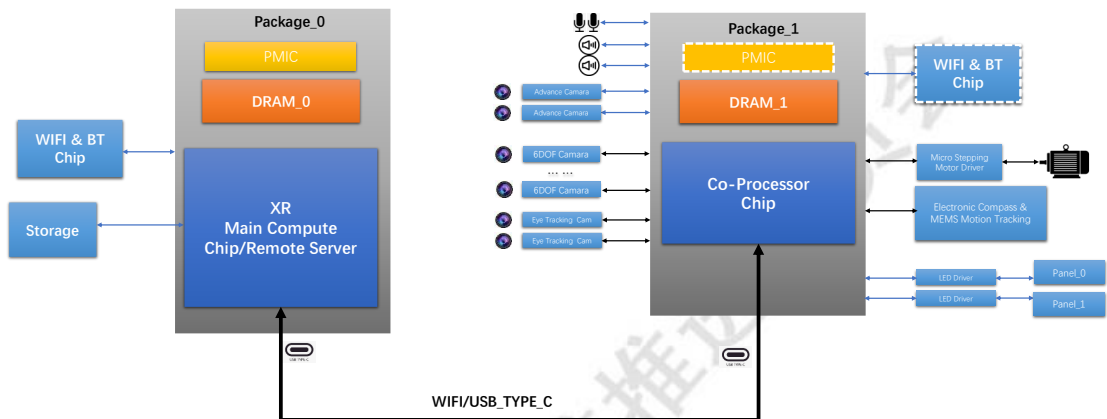


图 5.3 分体式 XR 芯片原型平台

受眼镜体积与电池限制，对于 AR 产品平台，更多采用分体式芯片平台设计，如上图所示，主控 SoC Die 和协处理器 Die 通过高速 I/O(USB 3. x/4)口，或通过 WI-FI 进行互联。分体式 (AR) 平台与一体式 (VR) 多芯片平台的整体主控 SoC 与协处理器功能划分相似，以下为差异部分：

协处理器 SOC 在分体式 (AR) 平台额外具有以下功能，其具有接收高分辨率摄像头的影像数据流，高像素图片捕捉的能力，提供完整多媒体支持，可完成视频编解码、音频编解码、Mic 数据接收、DAC 等音频服务，支持 DP Alt Mode 高速 USB 3. x/4 Type-C RX/TX 服务，

具有高性能 DPU 来支持单眼 4K，双眼 4K x2 级别的高帧率影像输出，且具有 ATW，ASW 帧补偿，可变帧率等功能，具有高速接口可通过连接外置芯片与主控 SOC 通信，交互数据和任务。对于 WI-FI 连接方式需具备独立的 PMIC 芯片与电池管理，即 Co-processor 内部具有较强功能 PMU 实现完整电源管理服务。对于 USB 3. x/4 type_C 连接方式的协处理器，在较低功耗需求的情况下，可无需 PMIC，通过 USB 直接供电。

5.3 现有扩展现实设备主控芯片

5.3.1 Qualcomm XR/AR 平台

高通公司深耕 XR 领域，拥有 Snapdragon AR2 平台、Snapdragon XR2+平台、XR2 平台、XR1 平台。



AR2 平台专为分布式处理 AR 打造，优化尺寸、功耗、性能，可支持 Wi-Fi7，支持定位追踪、环境感知、平面检测等功能，采用 4nm 工艺制程，具有眼镜端分布式处理能力，其 AR 协处理器帮助摄像头聚合 AI 和计算机视觉，使得 AR 眼镜更轻薄。



高通骁龙 XR2+Gen1 平台，采用 7nm 制程工艺，8 核处理器，支持 AI 算力 15TOPS，8K@60fps&4K@120fps，7 路并发摄像头，支持定制双 ISP、See-through、眼球追踪、表情追踪、语音输入、局部渲染、地图

构建等算法。

5.3.2 Unisoc XR 平台

紫光展锐当前在电视、投影、商显等领域里大规模量产，凭借 8K 超高清、XR、元宇宙兴起的契机，努力构建 4K/8K 超高清智能显示及元宇宙芯片平台，融合虚实边界，承载产业生态。



W517 采用四核处理器，提供多样化 AI 场景能力，采用 EPOP 和超微高集成技术，全新双 ISP 设计，可以为双目全彩阵列光波导、SLAM 空间定位提供强大基础能力。W517 芯片采用变速调节系统调度技术，对 SOC、调制解调器、操作系统、存储四个维度进行全方位功耗优化。



6870W 芯片采用 8 核处理器，先进的 6nm EUV 制程，是面向高端智能显示的芯片平台。融合新一代 AI 专用加速引擎的 4 核 ISP 架构，采用最新 Vivimagic6.0 影像引擎，全通路 4K HDR+ 显示引擎。支持 120 帧高刷新率，支持双屏异显，多屏异显，并支持 8 TOPS NPU 算力，最高支持 6 路 camera，适配多种外部 WI-FI6 方案。

5.3.3 Rockchip XR 平台



瑞芯微电子拥有多款 XR 旗舰级芯片。瑞芯微电子 RK3588 平台，采用 8nm 制程和 Arm 四核 Cortex®-A76

+ 四核 Cortex®-A55 的大小核架构，主频最高 2.4GHz，支持 8K60fps 解码能力，最大支持 6 路 MIPI CSI 摄像头输入，48M ISP 像素处理能力和 6T AI 算力，支持国密安全能力和安全启动等特性。

扩展现实产业及标准推进委员会
(TC625)

第六章 扩展现实操作系统

只有芯片，没有相应的软件支持，整个系统仍然是无法运行的，必须通过操作系统实现对芯片及外设各模块的管理，并对应用开发者提供友好的开发工具和底层运行库才能保证硬件能力能发挥出来，为各类应用程序的运行提供坚实基础。

6.1 扩展现实操作系统简介

广义上讲，XR 操作系统（简称 XR OS）就是在狭义的操作系统（如 Android、Linux 或 Windows）之上，面向 XR 设备打造的专用操作系统平台。该操作系统除了具备传统操作系统并发、共享、异步、虚拟等特性之外，还具备低延时、高实时（定位算法、手势交互等多种算法实时计算）、3D 多窗口交互等多种特性，并且基于 OpenXR 等互联互通标准实现 XR 核心功能，为 XR 设备提供标准化接口，降低开发难度、减少碎片化，加速 XR 生态建设以及提升产品体验。以基于 Android 打造的 XR OS 平台为例，为了达到低延时特性，XR 算法并不通过传统的 Android 的驱动层、HAL 层、Framework 层到应用层，获取数据输入给算法，而是从底层驱动直接把数据送给算法，大幅降低感知延迟，同时会充分利用芯片的硬件加速接口来实现多种交互算法的更高实时性。

目前全球 AR、VR、MR 眼镜厂商众多，一般都会基于 Android、Linux、Windows 专门为自家设备打造一套专用 XR OS，并提供相应 SDK、编

译器、模拟器等配套工具，便于开发者开发。综合不同设备的特点，XR OS 平台框架如下：



图 6.1 XR OS 平台框架

最底层是基础OS层，一般主流XR OS都是基于Android、Linux和Windows打造，不重复造轮子，复用当前已有的基础操作系统能力，进行针对性的优化与实现，可以大大减少人力投入和开发工作量，提升产品进度和稳定性。

往上一层是XR核心能力层，这是不同于以往平面显示的、基于键鼠和触屏交互的二维操作系统最核心的部分，一般以服务、引擎或者SDK的方式提供。这里也分为三个部分：感知系统、理解系统和展现系统。

感知系统包括声学感知、图像感知、运动感知和深度感知模块。XR 设备的声学感知部分包括 2-6 个不同朝向不同位置组成的麦克风阵列，可以收集多通道音频数据，DSP 模块可以基于此做降噪、寻向等声学信号处理。图像感知部分是 XR 设备重要模块，既有环境感知和头部追踪的可见光摄像头，追踪面部和眼球的红外摄像头，也有可感知深度的 TOF (Time of Flight) 及结构光摄像头。运动感知部分，主要指惯性动量单元 IMU，以近 1KHz 的采样率采集运动数据。

理解系统，既包括对人的理解，比如语音识别、眼球追踪、裸手追踪、手柄追踪以及头部运动的追踪等，也包括对环境的理解，比如基于 CV 的 6DoF 追踪、基于三维重建技术的实时环境映射、基于图像识别的二维图像追踪、基于物体识别的三维对象跟踪等。这里需要各种前沿视觉算法技术、硬件加速、机器学习等多个专业领域技术储备和工程化经验。

展现系统部分，不同类型的 XR 设备有些许差别，主流 VR 眼镜主要采用菲涅尔透镜和 Pancake 光路折叠两种光学方案，佩戴者无法直接看到现实世界，只能通过相机的视角以 VST (video see-through) 的方式观察外部环境。而 AR 眼镜目前主流的光学方案有 birdbath 和光波导等，可以直接通过眼睛以 OST (optical see-through) 方式观察外部环境。因为光学方案差异，VR 的 MR 实现方式以 VST 路径发展，追求沉浸式体验，未来通过 RGB 相机渲染出尽可能逼真的现实三维环境，AR 的 MR 实现方式以 OST 路径发展，追求现实增强体验，未

来重点实现虚拟渲染场景与外部真实环境融为一体、真假难辨的逼真体验。随着光学显示视场角（简称 FOV）增加，在 XR OS 里实现像素级别畸变矫正和色差矫正也是非常必要的。当然要想获得舒适的 MR 体验，操作系统层面实现 ATW 和 PTW(Positional Timewarp) 等低延迟渲染技术，尤为重要。

XR OS 的框架层，实现一些核心服务，XR UX 交互服务辅助开发者快捷开发出满足 XR 通用交互设计的 XR 应用。多窗口管理服务支持数以百万计的 2D 应用在 3D 空间中进行多窗口呈现以及交互。动态资源管理，满足一些云端编辑模型资源展示的需求，在 XR 上可以零代码呈现。OpenXR Runtime 支持 OpenXR 标准，支持开发者跨设备移植开发应用，降低开发门槛，提升开发效率。

XR OS 的开发工具层，主要指给开发者提供能力丰富、平台多样的 SDK，涉及 Android、Unity、Unreal、WebXR 等多个平台的支持。

XR 应用层，包含系统应用，开发者应用以及合作伙伴的行业应用等。系统应用，一般有：

- Home，基于渲染引擎开发的 3D 场景，Launcher 主界面；
- Gallery，图片浏览器，支持 2D/3D/360 度图片展示；
- Video Player，播放器，可播放 2D/3D/360 度视频；
- Cinema，影院模式，在电影院场景下播放 3D 视频；

- Setting, XR 系统功能设置 , 包括 BT/WI-FI/时间/开发者模式 /主题等;
- Browser, XR 浏览器;
- Account, 账号注册/登入/管理;
- APP Store, XR 应用商店;
- Input Method, XR 输入法, 包括英文/中文输入;
- Games, 一些系统内置的 XR 游戏应用;

6.2 扩展现实核心模块

6.2.1 6DoF 头部追踪

6DoF 头部追踪功能主要通过双目或四目鱼眼摄像头实现了头部的运动数据的生成, 包括 X、Y、Z 方向的位移数据以及 pitch、yaw、roll 的旋转数据, 通过该数据可以实现在 VR/AR 虚拟场景中移动。

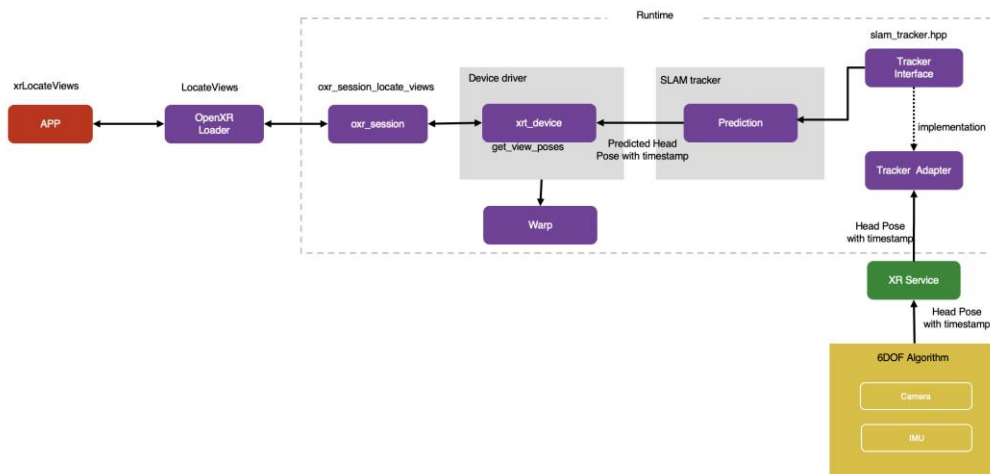


图 6.2 基于 OpenXR 的头部追踪框架图

在 OpenXR 的框架中定义了 6DoF API，通过 OpenXR Runtime 来实现头部 6DoF 数据调用，通过 XR Service 获取相应的 6DoF 数据，各个厂商可以通过实现 6DoF Algorithm 模块来适配自己的算法。

6.2.2 3/6DoF 手柄

3/6DoF 手柄功能主要实现了与虚拟场景交互，包括选择、确认、拖动、旋转等基本交互功能，在 OpenXR 框架中定制了 3DoF/6DoF 手柄 API，通过 OpenXR Runtime 来实现头部手柄数据调用，通过 Controller Service 获取相应的手柄数据及设备管理，其中 XR Input Service 以及 Input device 属于插件，由各个厂商实现自己的 3DoF/6DoF 手柄算法以及驱动。

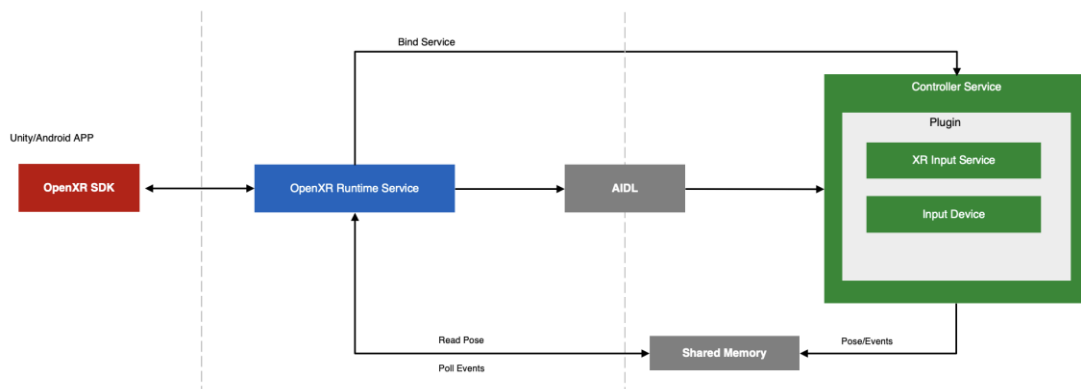


图 6.3 基于 OpenXR 手柄追踪框架图

6.2.3 Video See-through

Video See-through (VST) 是一种视频透视技术，主要用在 VR 设备上。从用户的角度来考虑，可以把 VST 定义为“让用户通过 Camera

捕捉的画面来观察虚实叠加后的场景”，VR 上通常使用 2 个 Mono 或者 RGB Camera。VST 实现原理就是通过 VR 头显的 Mono 或者 RGB Camera 捕捉到真实空间中的实时画面，然后将头显中的虚拟世界和 Camera 拍摄到的真实世界的影像进行叠加融合，最终在屏幕上呈现虚实叠加的效果。VST 的核心模块如下：

Camera HAL 模块，主要实现左右 camera 数据读取、图像效果优化等功能

XRService 模块，主要实现 Camera 帧管理、帧同步以及获取校准相关参数

Compositor 模块，主要用于 Camera 画面以及渲染画面的合成，在接收到 Camera Frame 后，需要对即将参与合成的两个 Camera Frame 在时间上对齐，同时也根据标定得到的外参对 Frame 加以空间对齐处理，最终将 Camera 图像和其他 Layer 混合叠加送给 Display 显示。

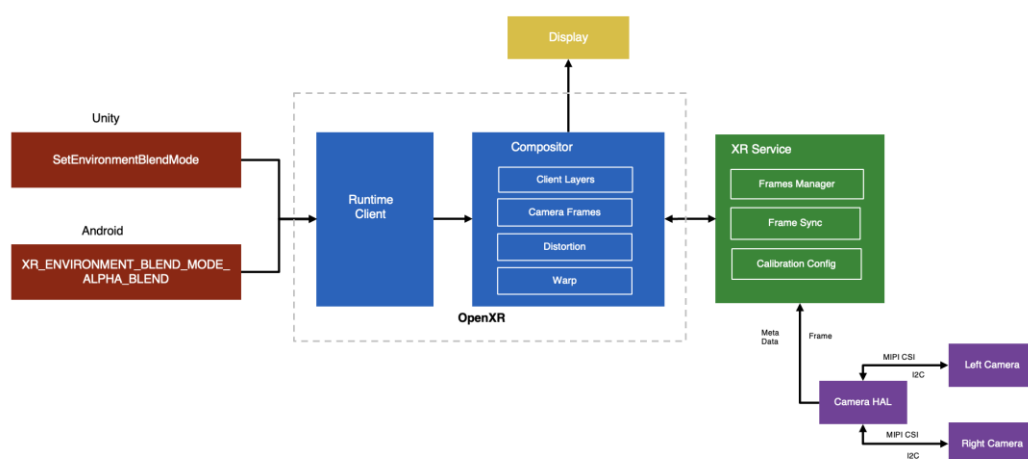


图 6.4 基于 OpenXR 的 VST 框架图

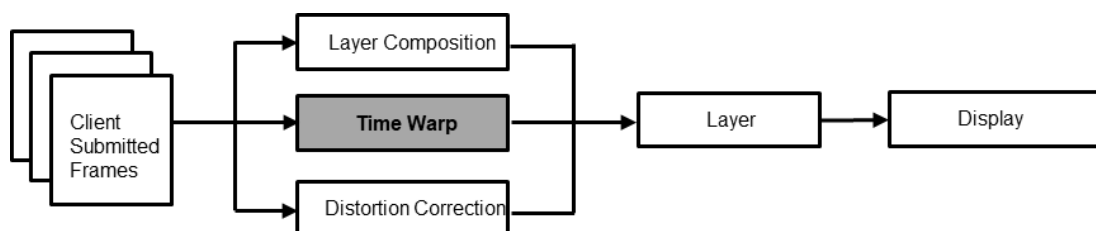
6.2.4 ATW

ATW 技术就是通过计算出帧渲染和上屏之间的用户姿态差异，在将渲染帧提交给 Display 之前，结合当前的头部姿态完成一次渲染，从而减小帧渲染和上屏的姿态差异。

ATW 是一个独立于渲染引擎的线程，线程优先级通常比较高，以便于在渲染引擎没有产生新的帧时，通过 ATW 技术插入一帧弥补当前丢帧的情况。

在 OpenXR 中，完成 ATW 必须要计算出一个合适的 WarpMatrix 矩阵，并将该矩阵作用于纹理，实现对纹理坐标的移动。此部分的工作将会在 OpenXR Runtime 中完成，当 Warp 线程被唤醒，它会接收新的 Pose 并作用于左右眼。

Warp 的工作将在 Compositor 中完成，当线程唤醒并接收新的姿势时，它会为双眼 Warp 相同的姿势。这是因为左右眼的图像将自上而下同时渲染，所以双眼不会单独进行 Warp。Warp 将首先作用于图像的上半部分并将其发送到显示器。接下来，它跳到图像的下半部分并扭曲图像的剩余部分。因此为了 Warp 的效果，显示屏的光栅顺序和 Vsync 时间和周期将尤其重要。



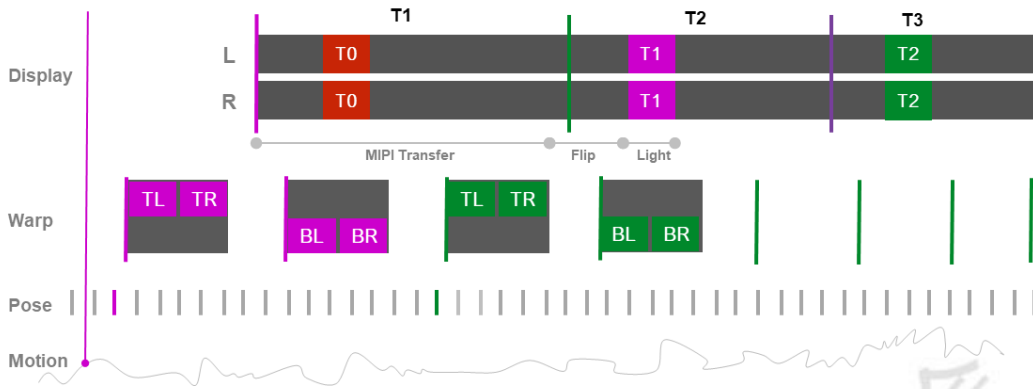


图 6.5 基于 OpenXR 的 ATW 时序图

6.2.5 畸变矫正和色差矫正

在 OpenXR 中，畸变矫正的工作处理同样在 Compositor 中完成。完成畸变矫正的方法有很多种，在 OpenXR 中采用 "Three approaches to VR lens distortion" 中提到的第二种方案 "Mesh based solution"，Distortion Mesh 是一种基于三角网格 (例如光学厂商给出 20 行*20 列的网格) 的解决方案，假设屏幕单眼的分辨率是 2280*2280，我们不需要处理屏幕的每个像素点，只需计算这个三角网格的 20*20=400 个定点，这个方案下 GPU 在处理纹理时会做相对合理的插值计算，因此可以节省 GPU 许多直接的运算。

需要光学厂商提供左右眼 Mesh 文件，Mesh 文件描述了原始网格点坐标和网格点变形后对应的点坐标。同时，为了完成 CAC 的调整，畸变校正后的坐标需要给出 R、G、B 三个分量的值。

畸变矫正 (Distortion Correction) 和 CAC 的工作在 OpenXR 中会统一用一个 Shader 完成。



图 6.6 基于 OpenXR 的畸变矫正示意图

6.2.6 无线渲染

无线渲染方案主要用作 VR 一体机游戏内容的扩充，通过该功能可以实现在 VR 一体机上串流 PC 主机的游戏，如下为软件架构图：

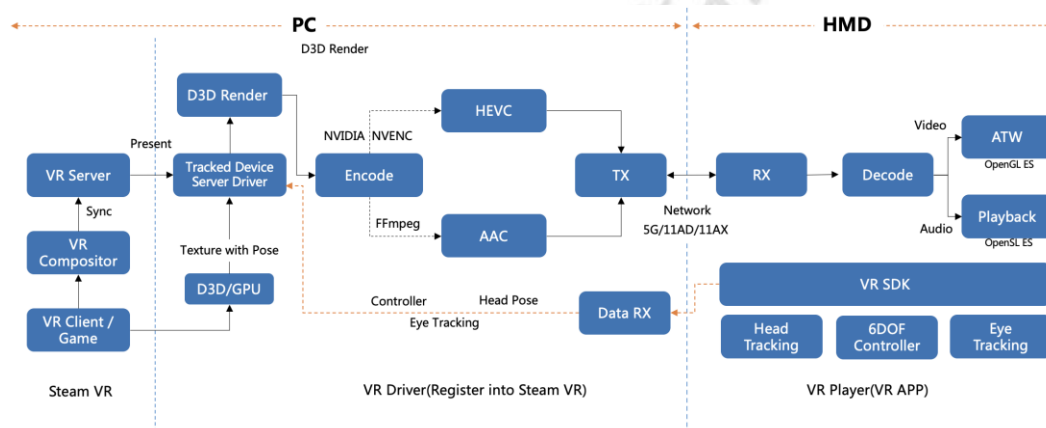


图 6.7 无线串流软件框架图

目前无线串流方案包含了 PC 端串的 Server 以及 HMD 端的 APP

- PC 端包括 2 个模块
 - SteamVR，主要用于接收 Tracking 数据并输出相应的渲染画面。

- VR Driver，主要用于对渲染画面进行编码、Tracking 数据的接收、解析以及无线传输。

- HMD 端包括 1 个模块

- VR Player，该模块会通过 VR SDK 获取到所有的 tracking 数据并打包，通过无线传输模块发送给 PC；同时通过无线传输模块接收到的视频数据进行解码并进行显示。

6.2.7 注视点渲染

注视点渲染功能需要结合眼球追踪摄像头模组以及相应的算法，才能够实时捕捉眼睛视线的位置，并提供给到 GPU，进行动态的分辨率调整。

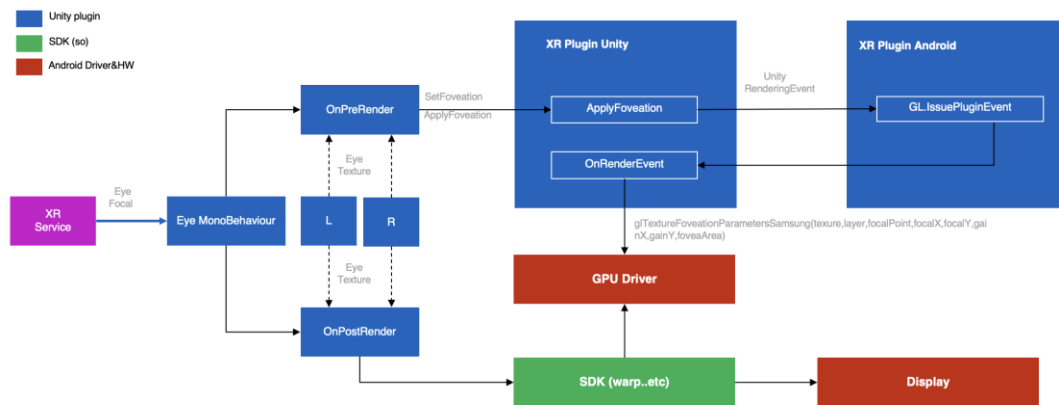


图 6.8 注视点渲染软件框架图

6.3 OpenXR 和 GSXR

6.3.1 OpenXR 支持

OpenXR 作为一种跨平台的标准，由开放标准行业协会 Khronos Group 制定，有助于打通各个平台，构建多平台互通的桥梁，允许游戏和其他应用程序在各种硬件平台上轻松运行，无需额外适配工作。Open XR 作为 XR 领域的一个重要里程碑，已经得到包括 Meta、Unity、高通、YVR、Pico、中科创达、Rokid 等国内外厂商的支持。

OpenXR 的系统架构设计如下图：

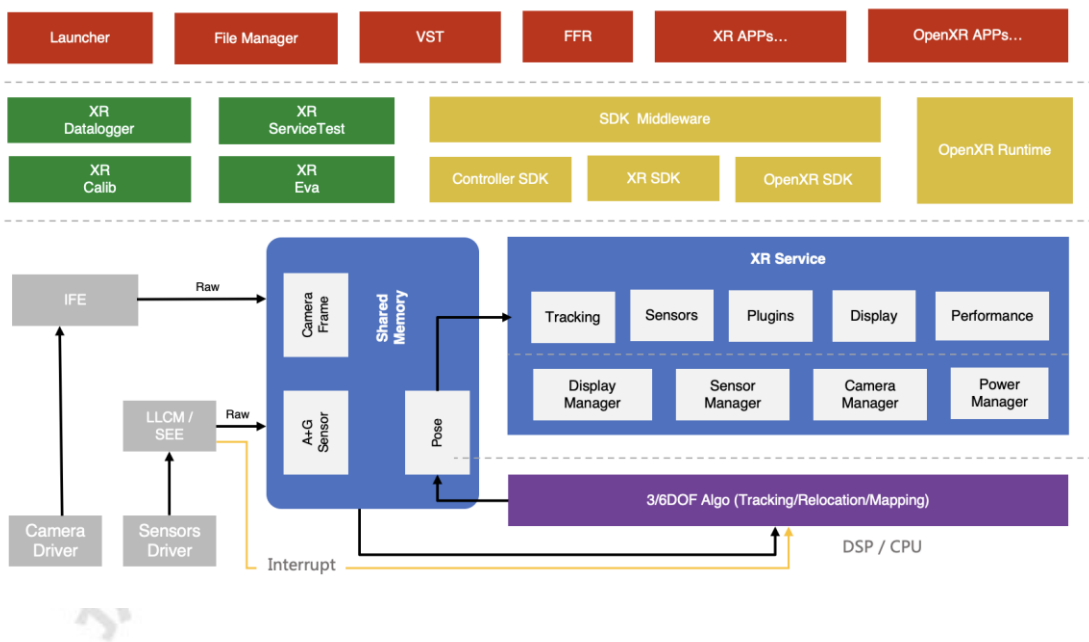


图 6.9 OpenXR 的系统架构图

OpenXR 的核心组成部件如下，

-
- OpenXR SDK 为 AR/VR APP 提供了一组功能，运用这些功能可以通过 API 的调用完成。
 - OpenXR Loader，属于 APP 的一部分，APP 在使用 OpenXR API 时会调用到 Loader 的接口，Loader 会将 API 分发到 OpenXR Runtime 中 (API 的实现在 Runtime 中完成，如果有 API Layer 存在的话，APP 发起的 API 调用会先到达 Layer 层进行处理，最终会转发到 Runtime 中)。
 - OpenXR Runtime 由 Runtime Client、Runtime Server 和 Compositor 组成。这三个库将会一起打包进 `openxr_runtime.apk` 中。OpenXR Runtime 作为 OpenXR SDK 的功能实现，当 APP 运行时，它将从 `openxr_runtime.apk` 加载 OpenXR Runtime Client 的共享库 Runtime 进行交互，Runtime 完成对 APP 提交的显示 Buffer 的处理、TimeWarp、Distortion Correction、CAC、Pose Prediction 以及 Controller 等重要功能的实现，最终由 Compositor 将渲染合成的画面提交给 Display 显示。
 - XR Service 整合设备的硬件模块和算法，支撑 OpenXR Runtime 核心功能的正常运行。OpenXR Runtime Server 和 Compositor 可以从 XR Service 中获取到 6DoF 数据和 Camera 的数据完成 APP 所需的功能。

6.3.2 GSXR 支持

General Standard for XR（简称 GSXR）标准通过提供统一的应用开发标准和设备对接规范，帮助开发者实现快速分发和多平台覆盖，更好地解决硬件终端平台耦合、标准差异化严重等问题，增强软件适配性。目前已有多家 XR 厂商完成了适配，如 PICO、NOLO、创维、爱奇艺、Rokid、Nreal 等。

GSXR 是由 GSXR SDK、GSXR API、GSXR Loader、GSXR FrameWork、GSXR Runtime 等部分构成。GSXR 系统架构如下图，目前 40% XR 厂商采用图 1 的方式，60% XR 厂商采用图 2 方式，针对两种方式，GSXR 都会提供相应的参考 Demo 以及源代码。

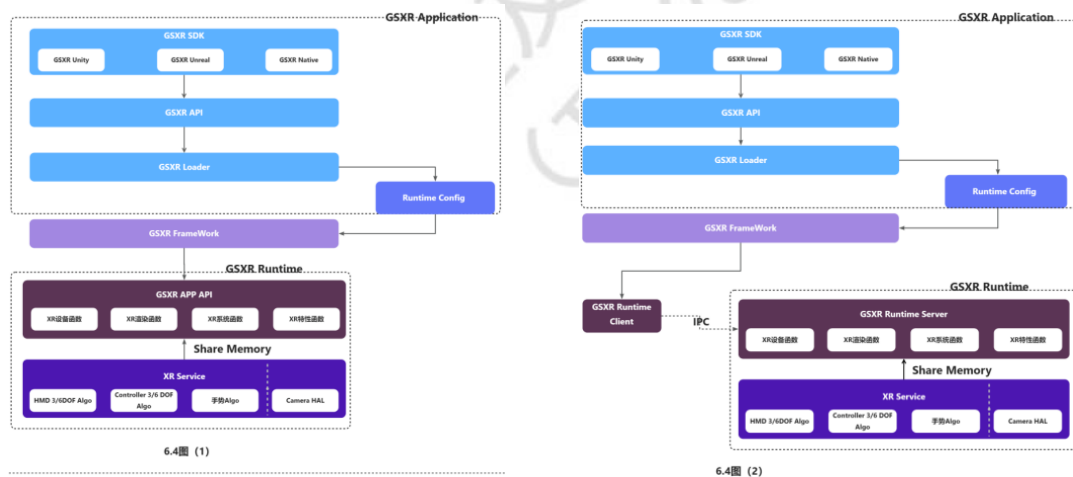


图 6.10 GSXR 的系统架构图

GSXR SDK 目前已提供 Unity、Unreal 两种引擎使用的 GSXR SDK，GSXR 支持 Unity 5.6 及以上所有版本，在 Unity 引擎下提供 Unity XR 版本的 GSXR SDK 及 GSXR Legacy 版本的 SDK。Unreal 引擎提供了

4.24、4.25、4.26、4.27 及 Unreal 5 版本的 SDK。不同版本的 SDK 均提供相应的快速入门文档。

在 SDK 基础上，还提供了重要扩展性插件，如 GSXR 搜狗输入法插件等。同时还为开发者提供了多种 XR 案例，如手势案例、音视频案例、物理特性案例、全景音案例、瞬移案例等，以便开发者更快进行开发。

扩展现实产业及标准推进委员会
(TC625)

第七章 未来扩展现实技术展望

目前 VR 设备形态和技术基本稳定，趋于统一，但 AR 设备因为使用场景、技术实现的差别，还无法统一。

面向工业场景的 AR 设备，比如微软 HoloLens 系列、Rokid X-Craft 系列等，在近 3 年时间内还会围绕高性能、高精度定位、续航时间长的 AR 一体机方向迭代发展。

面向消费者观看场景的 AR 设备，比如 Rokid Air 及 Max 系列、Nreal Air 及 Light 系列、雷鸟 Air 系列等，定位支持 3DoF 及 6DoF 跟踪、头手交互等场景。未来其眼镜形态可能往无线 AR 分体机或 AR 一体机方向迭代发展，可满足 2-4 小时续航时长的观影娱乐、文旅导览等场景使用需求。面向信息提示场景的 OPPO Air Glass 系列、Rokid Vision 系列等，未来会继续发展成以轻算力、低功耗级芯片方案，结合 Micro-LED 光波导显示技术，打磨成可以满足全天使用、Always-On 的极致轻便 AR 一体机，预计也将很快成熟。

随着算力提升，未来的扩展现实设备将会需要更大带宽，更低延时，更高沉浸感，更强的交互能力。

7.1 无线传输技术

在无线传输方面，Wi-Fi 6E/Wi-Fi 7 将成为标配，5G/6G 也将给用户带来全新体验。

7.1.1 Wi-Fi 技术

最新的 Wi-Fi 技术 802.11be 也就是 Wi-Fi 7 技术支持高达 320MHz 的信道。最大支持 30Gbps 的吞吐量，支持 16 空间流。即使 XR 因为空间所限只能实现两条（2x2）空间流，其物理层性能也能达到 4.6Gbps，完全可以支持一般的 XR 应用需求。时延方面，最低时延预期低于 5ms 是 Wi-Fi7 工作组的目标。相对于 Wi-Fi6 的 20ms 延时性能，Wi-Fi 7 实现了巨大的飞跃。同时 Wi-Fi 6E 以及 Wi-Fi7 引入了 6GHz (5.925GHz~7.125GHz) 频段，同时支持 2.4GHz、5GHz 和 6GHz。在新频段上可以有效避免老旧设备的干扰，从而提升相应的网络性能，保证了 XR 的用户能够得到高带宽、低时延的特性网络。此外 Wi-Fi7 还支持很多新特性。总之，未来的 Wi-Fi 7 技术将在 XR 领域发挥重大作用。

7.1.2 5G/6G 技术

5G 与 6G 是 XR 未来无线通信的发展的主要方向。XR 作为新兴业务，要求在支持低时延高可靠的同时保证大带宽，这对 5G 网络提出了新的挑战。5G 采用边缘计算让云端的计算、存储能力和内容更接近用户侧，使网络时延更低，用户体验更极致，使能 AR、VR 和 MR 等应用。同时，得益于 5G 低时延、大带宽能力，终端侧的计算能力还可以上移到边缘云，使得 VR 头盔等终端更轻量化、更省电、更低成本。这种“轻终端+宽管道+边缘云”的模式将砍掉昂贵的终端的门槛，摆脱有线的束缚，从而推动 XR 应用普及。云化 XR 系统将实现用户和

环境的语音交互、手势交互、头部交互、眼球交互等复杂业务。在 6G 技术的支持下，基于云化 XR 的总时延预计低于 10ms。

7.2 多媒体音视频技术

沉浸式对多媒体处理提出更高要求。限于产业实际，现在的视频沉浸式体验还是通过“H. 265/H. 264 视频编码+360 度空间投射”的方式提供 3DoF 的内容体验。对于 VR 来说，下一代视频编码技术要能支持真正的 6DoF 沉浸式体验，让沉浸式体验效果更好。

和手机/耳机广泛应用的空间音频(剧场体验)比较，扩展现实增加了更多维度。前者侧重于声源在一个预定义的与现实无关的空间的表达，听者在实时聆听中具备 3DoF 的方向调整。而后者需要声源在一个可变互动的空间的表达，同时听者具备 6DoF 的调整和现实空间映射。例如，混合现实中的虚拟乐器需要恰到好处地表达当前现实空间和位置的反射和混响特征，虚拟会议室中需要把远端的会话去混响后再按听者所在环境的特征进行渲染回放等。基于上述需求，多家厂商和组织，如 Meta、微软、MPEG、Rokid 在技术和标准上有多领域的预研：声源体积和指向性渲染、实时声线跟踪渲染、现实空间的盲估分析与合成、头戴设备多麦声场采集、多麦采集声场融合、个性化采集无感化快速化等。

7.3 智能交互技术

7.3.1 手势追踪技术

手势追踪技术在 XR 交互领域会起到非常重要的作用，手作为身体的一部分，可以提供更自然、更直观的交互方式。但是手势跟踪作为依赖视觉跟踪和识别的技术，也会受到一定的挑战：

- 1) 精度受限：手势追踪技术已经相当成熟，但在某些情况下可能精度受限或者识别错误，例如识别复杂有遮挡手势动作等。
- 2) 环境依赖：手势追踪技术需要适当的光线条件，以确保手势能够准确被捕捉和解释。这是在室内环境下使用这种技术的优势，但在室外或恶劣的光线环境下，手势追踪可能存在挑战。
- 3) 可访问性挑战：手势追踪技术对于某些人（例如身体残疾、运动能力受限）可能存在挑战，因为需要进行特定的手势动作或者不完全依赖手势可以闭环交互来控制应用程序。在开发应用时需要考虑到可访问性问题，以确保所有人都能够享受到 XR 交互带来的优势。

当前主流的基于视觉识别的手势追踪有一定的局限性，所以手势追踪技术配合其他传感器和交互技术使用可以提供更全面的 XR 体验，以及更高效的交互方式，也是未来 XR 行业挺好的发展方向。

-
- 1) 眼动追踪相结合：可结合头部追踪技术和焦点注视来实现更加自然的交互方式，当用户眼睛注视到一个物体时，系统可以自动感知到这个操作并响应对应的操作，这样可以避免频繁举手产生的交互疲劳。
 - 2) 指环交互相结合：在虚拟现实场景中，指环交互则可以通过震动和光效反馈等方式来增强用户手势交互沉浸感，提高交互的真实感和趣味性。同时指环上的传感器，可以一定程度上补充手部部分遮挡场景的识别和跟踪体验。
 - 3) 深度传感器相结合：与当前的一些 ITOF (Indirect Time of Flight) 相机不同，未来 XR 设备会结合 DTOF (Direct Time of Flight) 深度相机使用，DTOF 则是一种通过反射回来的光线的时间差来计算物体与传感器之间的距离的技术，可以提供更高的精度、更低的功耗、更低的算力，还可以解决传统摄像头在低光环境下无法进行手势追踪的问题，可以大大扩展手势交互的使用场景。

未来，随着芯片技术的发展、手势算法的成熟，新型传感器的普及，手势追踪技术在扩展现实领域的价值非常大，可能成为完全替代手柄交互技术的一个非常有竞争力的选项。

7.3.2 眼动追踪技术

当前，眼动追踪技术还有很多难点需要攻克，比如，眼动算法能更准确地基于所采集的原始眼动行为来“认知”用户意图，减少用户个体和环境差异（眼球角膜、佩戴眼镜、周围光线等）对系统追踪精度和通用性的影响，不用完成眼动追踪标定校准就能直接使用，方便用户操作。

7.3.3 触觉反馈技术

触觉是人的重要感官，也是 XR 体验的未来方向之一。触觉反馈是使用技术来刺激触觉，再现用户直接与物理对象交互时的感觉。这些效果可以从简单到复杂，简单的如游戏角色被击中时手机上的振动；复杂效果如游戏中试图推门时，PS5 控制器触发器的自适应阻力和触觉手套的逼真纹理感觉。触觉技术给 XR 设备提供更沉浸式的用户体验，并融合听觉体验让用户身临其境。

触觉体验目前是行业科技巨头的重要探索方向，行业很多公司在 XR 配件上已布局，发布了相关产品，其中最前沿的是触觉手套和背心。多家触觉技术公司在此领域展开研究和竞争，从低端的手和身体的跟踪和振动效果，到完全逼真的力反馈和触觉微流体效果，以及 Teslasuit 的“全身触觉动捕服”。2022 年，Meta 收购了德国初创公司 Lofelt，该公司专门从事触觉技术，通过智能手机或视频游戏控制器等硬件设备中的力或振动在复制虚拟环境中的触摸体验。另

外，新触觉设计前沿初创企业 Hap2U 和 Aito，它们将 2D 触觉体验引入屏幕和触摸板，为 XR 触觉体验创新提供了一个新方向。Hap2U 在屏幕边缘使用压电超声波感应器，可以在屏幕任何点上产生高分辨率触觉效果。这种将触觉集成到屏幕中是一种技术趋势，该技术仍处于起步阶段。随着屏幕制造技术的进步，屏幕单个像素的大小随着时间的推移将不断缩小，这意味着活动显示区域仅是显示器整个区域的一部分，还有更多空间可以添加各种新功能。

7.4 环境理解技术

环境理解是指系统能够理解和感知现实世界的环境，包括空间结构、场景特征、物体属性等，同时也能够对场景进行实时的分析和推断。通过环境理解，XR 系统可以更加准确地感知用户的周围环境，实现更加自然的交互方式，对平面的理解（如墙面、地面、桌面）可以在平面上呈现模型交互，对物体的追踪和理解，可以实现真正的增强现实体验，对环境中纹理深度的理解，可以实现一些真实的碰撞和遮挡效果。

随着 AI 技术的发展，环境理解在扩展现实（XR）领域有望经历许多创新和变革。

- 1) **更高的 SLAM 精度：**AI 算法变革和芯片计算能力的提升将使得环境理解技术更加精确和实时。比如使用卷积神经网络（CNN）可以

快速识别关键点和特征。这将使 XR 体验更加自然、流畅和沉浸式，提高用户体验。

- 2) **更好的地图生成：**SLAM 算法最重要的任务是创建准确的地图。通过使用深度学习技术，识别复杂场景中的物体和结构，如遮挡物、反射和半透明表面，可以提高地图的准确性和可靠性。
- 3) **更优的自主性：**SLAM 算法可以通过深度学习技术自主学习环境（室内外环境、光照条件、物体形状和质地等）并改进自己的性能，而无需人为地进行修改。这将使算法更具广泛适用性和可靠性。
- 4) **更准的语义理解：**借助 AI 技术，XR 系统将能够识别和理解场景中物体的语义信息，如物体类别、属性和关系。这将使 XR 应用能够根据上下文提供更智能、个性化的交互和服务。
- 5) **更强的动态追踪和预测：**AI 技术可以帮助 XR 系统更准确地追踪和预测场景中物体和人的动态变化，提供更稳定、流畅的 XR 体验。
- 6) **多模态融合：**AI 技术可以实现多种传感器数据的融合，如视觉、深度、惯性传感器、移动 5.5G 高精定位等，提高环境理解的准确性和鲁棒性。
- 7) **更高效的算法和硬件：**随着 AI 技术的优化和专用芯片的出现，比如使用专用硬件加速器——如 AI 芯片和 GPU——提高计算性能和能效比，算法将在较低的计算和能源消耗下实现高性能，为便携式 XR 设备带来更长的使用时间和更好的性能。

7.5 生物电感知交互技术

生物电感知交互技术指的是通过感知和解读人类脑电（Electroencephalogram，简称 EEG）和肌电（Electromyography，简称 EMG）等生物电信号，实现人机之间自然、高效的交互的技术。在 XR 领域，生物电感知交互技术可以帮助人们更自然地与虚拟环境进行交互。根据目前的一些应用场景和研究方向，本白皮书着重介绍基于 EEG 的脑机接口技术和基于 EMG 的肌电交互技术。

脑机接口（Brain-Computer Interface, BCI）技术是一种直接将脑电信号（EEG）转换成电信号或其他控制信号的技术。该技术可以让人们通过意念控制计算机或其他设备，以实现某些特定的交互操作，比如用户可以通过使用自己的脑电信号控制设备上的菜单选择、控制游戏角色或 3D 模型移动等。基于脑机接口的神经适应技术可以通过监测用户的生理反应并根据反馈进行调整，以提高用户在 XR 中的沉浸体验。例如，当用户感到压力或焦虑时，系统可以自动调整场景。但是目前脑机接口技术的发展还面临一些技术限制。首先，脑机接口技术需要准确地检测和解读人类大脑的信号，但是人类大脑信号的检测精度和解读准确性目前还存在较大的局限性。其次，脑机接口技术需要提供高速的数据传输和实时解析处理能力，以及可靠的数据存储和模型训练技术，而这些技术也还需要进一步完善和提高。当前虽有诸多限制，但因为脑机接口技术可以实现与人类大脑直接的交互和控制，未来可能会用于提升 XR 设备的交互体验。

肌电交互技术是一种利用人体肌肉电活动信号来实现与计算机交互的技术。在 XR 领域，肌电交互一般以手表、腕带、指环等设备形态存在，通过采集多通道高帧率的肌电信号和陀螺仪数据，在海量训练数据的基础上，结合深度学习的肌电分解识别算法，可以精确感知到用户的交互行为，进而应用在 XR 领域的微动手势控制、输入法、全手模拟等交互场景。用户对 XR 设备交互确定性要求非常高，但人体间个体差异大，肌电信号也会受到皮肤和毛发干扰，肌电信号的采集和准确识别是当前面临的较大挑战，商业化落地尚需时日。相信未来在海量数据集结合 Transformer 大模型训练技术的基础上，与高算力的 XR 芯片相配合，肌电交互技术在 XR 领域必定会得到进一步发展。

扩展现实产业及标准工作组
(TC625)

第八章 结束语

由以上梳理我们可知，当前 XR 芯片还处于发展初期，用户体验方面还有很大提升空间，通用芯片不能完全满足现有 XR 设备发展的需要，需要有针对 XR 特定需求的专用芯片，以提升用户体验，降低功耗，吸引更多用户使用 XR 设备，最终形成出货量和 XR 生态的良性循环。

我们可以预期，随着无线传输技术、大算力芯片处理能力和 AI 技术的发展，XR 领域将会迎来新一轮创新浪潮，为更多行业和应用场景提供更加沉浸、智能和自然的体验，也将深刻改变产业业态和人们的工作生活。我们也期待通过产业链上下游合作伙伴的共同努力，XR 产业能得到更快更好的发展。