谢谢主持人，每年都来参加世界智能网联大会，每年都能够看到业界许多的进展和变化，这非常好，我想我们所有的业内从业人士都会感到惊喜。超星未来是一家年轻的创业公司，我们成立一年多，我们专注于面向智能驾驶提供计算平台的软硬件解决方案，所以今天的主题也是面向智能驾驶快速变革的计算平台。

先从几个热门的话题谈起，整车的电子架构向集中式演进。软件定义汽车，SDV的概念，未来的智能驾驶汽车的价值和个性化、差异化将会主要体现在它的软件，特别是智能传播以及其所延伸的技术服务上。所有的主机厂他们都在不遗余力地打造配备智能驾驶功能的量产车型，尤其是像最近这个话题非常火的就是特斯拉为代表的造车新势力的崛起，可能大家都会注意到在资本市场特斯拉的汽车的产销量仅仅是老牌车厂的领头，但是它的市值却已经超过了它们的一大截，这就反映了资本市场对未来和趋势的过程认可，这也必然会加速推进传统主机厂在这方面的变革。讲这些可能大家也都是耳熟能详的热点话题和趋势，从我们做计算平台的人，或者是公司的角度来看，看到什么？我们关注到这些话题和趋势的背后都跟我们的计算平台有着密切的关联。

大家都说软件定义汽车，如果没有计算平台，如何保证软件高效可靠的运行，而且计算平台就是指的软硬件的解决方案，软件本身也是计算平台的重要组成部分，刚才英伟达的同事也讲了，我们不是一个硬件公司。总结一个结论车载智能计算平台将会是新型汽车电子电气架构的核心，是自动驾驶智能汽车的“大脑”，在智能汽车量产落地推进的过程中间，它的重要性和价值会越来越突出，受到越来越多的关注。

这里介绍一下我们的智能驾驶计算平台的基础架构，首先从整个框图的右下角讲起，我们是有一个异构的硬件平台层，这里包含了通用的CPU处理器，MCU，然后还有性能处理核，以及多样的定制化的加速的IP，也是下面肯定要重点介绍的。硬件层上面需要一个系统软件层，包含了操作系统，析出软件以及实时运行环境，还需要工具软件支持上层算法在硬件上的优化和部署。也就是说我们说软硬件的计算平台解决方案就包括了异构的硬件，系统的软件以及工具软件。那么上层的应用模块和算法我们大家都能够知道的，自动驾驶域的感知决策、规划控制、定位、高精地图将会在计算平台上部署和运行。那么我们也根据基础架构的层次划分，我们也提出了计算平台的技术路线，总结起来就是我们左边的四句话，我们把它称之为硬件的异构化，软件标准化，工具自动化和应用模块化。开玩笑，我们国家智能汽车有新四化，我们也好好搞搞计算平台这个小的“四化”。

一个量产的计算平台仅仅有基础的框架以及技术路线是不够的，我们要考虑的问题非常多，甚至都不能仅仅考虑到算力。作为一个量产的计算平台的产品方案，还需要考虑成本，它的能效比，它的低延时、供销、灵活性、开放性以及产业生态。这里所面临的挑战是非常大的，现在我们能观察到智能驾驶以及计算平台这个行业内大家做到研究开发的今天，也已经在进行一些深层次的思考。也就是说平均的算法性能可保证的处理运行速度，这是个非常好的从系统的整体去评估性能的一个方法，我们也非常认同的。我们的方案是从整体上着眼提高系统的能效比、功耗、低延时方面的综合性能。

提到能效比这里有一个曲线图，上面有三条颜色不同的曲线，可能代表了不同的方法。我们先从蓝色的图线开始讲起，对通用的处理器，大家典型的看到的就是CPU，大家也都注意到了我们的CPU的发热量和功耗越来越大。它由于受到本身的半导体物理极限的限制以及工艺水平的极限的限制，它的能效比已经到了发展的瓶颈。如果放眼未来，方法三上面灰色的图线，可能有一些很好的想法，很新的器件，新的计算模型的方案，包括像量子计算比较热的这些，存算一体这些，它很有潜力，但是我们不知道什么时候可用，光计算这些很不清楚。

软硬件协同、定制化设计，就成为了我们提高系统的整体能效比的有效方式。为什么要在自动驾驶的平台方案里面我们要在异构的硬件上面做定制化加速呢？其实我这里用几个例子来说明这个问题。因为自动驾驶这个域里面包含的算法非常丰富多样的，不是只有神经网络，会有非神经网络对于算力需求仍然很高、很密集的场合，不能很好的处理，甚至会形成系统的延时的瓶颈。所以我们也举两个非常实际的算法的例子，像右上方的图表是一个激光雷达点云的感知算法，我们可以看到中间的柱状图里面，橙色部分是已经经过优化过后的神经网络处理的所代表的延时，它已经显著的低于了预处理部分，用CPU来处理的这部分的延时。而且这部分的计算不是能够简单的就用AI加速核去处理的，所以它实际上形成了一个系统的瓶颈。

我们再看一个案例，在个特征提取的算法上，我们还看到这么一个巨大的灰色的柱状图，由于缺少特定的加速可以说给CPU以及整个系统作为了非常大的延时压力。通过这个实际的案例我们可以看到，由于自动驾驶的域里面算法的广泛性、复杂性，仅仅有CPU和AI 的性能处理核是不够的，我们平台的架构就是包含了CPU，包括AI处理核，同时也包含了可灵活编程定制化加速的逻辑。

这张图表是基于FPGA的定制化加速IP点云预处理以及实际的测试，可以看到跟嵌入式的CPU相比，我们的FPGA的加速后的性能可以提升到5-10倍，即使是一个很先进的台式机的CPU相比，我们的性能也可以做到2-3倍的提升。在做到这么好的性能提升的同时，我们来看一看右下角的表格展示的用掉的硬件资源，仅仅相当于一颗ZU2或者ZU3的硬件资源，是非常技术领先和实用的地方。

特征提取的算法如何进行加速，我们做到的对比测试大概有20倍左右的性能提升，并且对于一个右侧的表格里面是一个高精的SPG，整体的资源占用里只有10%左右，我们的方案针对智能驾驶域里面主流的芯片无法高效处理非神经网络算力密集型的计算任务的情况下，我们的平台提供了这种定制化的加速的逻辑，并且我们都开发了定制化的IP，所以可以有效的广泛的支持自动驾驶全域的算法，图象感知的，激光雷达感知的，降低系统的延时，提高系统的整体的安全、可靠性和响应速度，提高系统的整体性能。

实际上异构的硬件平台，定制化的加速这条技术路线也已经成为业界企业的主流方向，英伟达都在采用这个技术路线。刚才主要讲到我们在硬件定制化加速方面所做的工作。这边会介绍一下软件协同优化来提升系统性能的图景。这张图表显示的是，在同等的工艺水平条件下，统一种类的芯片它代表的能效比的极限和性能是趋于一致的，受到本身的固有的限制。想要做到很好的能效比的提升，软件的协同优化是一个重要的途径和方式。基于这个需求，我们超星未来也开发了硬件友好的软件协同优化部署工具，也就是说我们的软件优化工具可以针对于CPU、GPU、FPGA以及其他种类的硬件对算法模型进行优化。可以针对硬件的不同特点，对不同的算法模型进行量化、压缩、减值这些操作。同时我们也支持不同算法模型的优化，我们也开发了基于NAS的模型结构搜索软件，谷歌结束的NAS的挑战赛里面超星的团队提交的方案也获得了第三名的成绩，这是国内的企业参与队伍里面最好的成绩。

这个是我们在做的Yolov3图象目标检测的优化效果，在3款不同的硬件上跑的性能提升的对比，包括了个英伟达的GPU，也包括了一款最右侧的不是很知名的国产的RK3399的芯片，我们看到它主要处理的性能每秒的处理能力都提升了大概3倍左右，而处理的算法模型的性能没有大的损失。再看高级别自动驾驶里面所必须用到的点云目标检测的算法的优化效果，两个主要的性能目标都没有损失，甚至Precision还提高了一个点，提高模型的泛化能力，避免过学习的存在。系统延时大概70毫秒左右，达到可用的水平。

刚才讲到的主要是硬件定制化的加速，以及软件协同优化，他们在提高整个系统比如能效比，低延恩时方面方法有帮助。关于计算平台所要的灵活性方面的考量，带来更好的异用性对上层的算法，首先要支持AI的开放框架，我们超星的平台本身就可以支持Caffe、PyTorch等同时可以支撑智能驾驶、环境感知、摄像头的目标检测、激光雷达的目标检测、实力分割，这些多样算法在平台上的优化部署。定制化的加速，不光是刚才讲到的点云的前处理，特征提取，包括基于视觉的高精定位以及点云的APP的匹配的算法，都可以利用超星平台的定制化的加速的IP进行部署。基于以上技术推出了超星未来面向了L3、L4级别的计算平台。我们的计算平台是一个软硬件的解决方案，在NOVA-30的基础之上，定制化的加速IP以及软件协同工具都可以一起来支撑L3、L4级别落地自动驾驶对计算平台的需求。

在产品落地上面我们已经参加了北京市科委支持的2022年冬奥会首钢园区的自动驾驶的示范模型，这个项目中与清华大学合作，为首钢园区L4自动驾驶运行提供计算平台以及整车的解决方案，参与的两辆车包括小鹏汽车的乘用车以及北汽福田的一辆商用火车。    
 最后简单介绍一下我们的公司和团队，超星未来是清华大学车辆学院与电子工程系跨院系联合创新成果落地的一个产物，我们的两位发起人，一位是车辆学院的首任院长杨殿阁教授，本身也是车辆交通旅行驾乘领域的专家，另外一位老师是电子系系主任汪玉教授，本人也是在人工智能系统领域有十多年的耕耘和积累，最近其实他也刚刚有一轮非常成功的创业经历，创始人团队主要来自于清华以及校友相关联的企业，也都是在这个智能驾驶、人工智能算法以及汽车产业有过多年的工作经验，所以我们也很高兴有这样一个机会，在智能网联大会上与大家交流，也希望业内的合作伙伴能够与我们联系。谢谢大家。