

镍价走高或影响新能源汽车电池技术方向



本报记者 姬晓婷

“我们刚刚进行了一轮涨价,现在不订的话,价格还会涨。”比亚迪北京销售网点工作人员这样告诉《中国电子报》记者。近日,比亚迪汽车迎来全面涨价。销售人员告诉记者,除销量最高的“汉”外,所有的汽车款型都进行了价格调整,其中EV系列涨价6000元,DM-i系列涨价3000元。包括比亚迪、小鹏、特斯拉在内的新能源汽车品牌均上调了其汽车价格。此轮新能源汽车涨价,很大程度上是受到镍价上涨的影响。镍价上涨,短期来看将带来新能源汽车相关产品价格上涨,长期来看将可能带来新能源电池行业方案的全产业调整。

镍价上涨	影响新能源电池行业	新能源汽车电池方案	或将出现变革
镍是电动车动力电池制造的上游关键原料,在动力电池中主要用于三元正极材料的制造。全球新能源汽车知名品牌特斯拉采用的三元锂动力电池,便需要大量镍金属作为原材料。据集邦咨询的数据,2021年全球镍矿产量约为270万吨,主要来自印度尼西亚、菲律宾和俄罗斯,其中俄罗斯镍矿产量占据全球总产量的约9%,位居第三名。受到俄乌局势的影响,镍价在国际期货市场出现剧烈波动。根据伦敦金属交易所(LME)展示的镍价历史交易数据,3月7日当天,镍价由3月4日的收盘价2.91万美元激增至5.03万美元,价格涨幅达137%。不过,近几日又出现了大幅回调。	镍价上涨已传导至新能源汽车电池和整车行业。国内某新能源电池生产厂商在接受《中国电子报》记者采访时表示,受到镍价上涨影响,他们正在与汽车企业商讨涨价事宜。当前电池产品开启全线涨价模式,涨幅在20%~50%不等,涨价电池类型包含对镍金属需求最多的三元锂电池,也包括铁锂产品。	“由于公司有原材料备货,故此次涨价对公司电池生产的影响不大。”该公司负责人表示。	2~3周即可提车。而当3月21日记者再次向该销售网点了解时,热门车型ModelY后轮驱动版交付周期已经延长至10~14周了。也有的汽车品牌交付周期受原材料价格波动影响不明显,记者日前从比亚迪某北京销售网点得到消息,除个别车型外,许多畅销车型仍有现车,最快3天内实现交付。
			芯谋研究分析师温旭表示,此轮镍价大涨受到俄乌局势影响,但这几天正在缓解。从3月7日以来的伦镍价格来看,自3月7日、8日价格达到高峰以来,镍价正在缓慢回落。镍金属价格已经从3月7日5.03万美元每金属吨以上逐渐回落至3.69万美元每金属吨。温旭表示,镍价回落是由于受到近期市场监管和恐慌抛售的影响,在他看来,镍价将逐渐下落,将有望下降到3万美元以内。
			镍金属涨价给采用三元锂电池的电动汽车厂家带来了很大的压力。温旭表示,一辆车所用的电池价值约占到汽车总价的40%。从市场上常见的电动汽车技术方案来看,磷酸铁锂使用的镍量比三元锂更少,且其技术发展更加成熟。在镍价走高的情况下,不排除原本采用三元锂技术方案的企业转为采用磷酸铁锂技术的可能。由此,镍价走高带来的将不仅是新能源汽车电池及整车涨价,将有可能带来新能源电池方案的市场变革。

东芝将投资8.3亿美元扩产功率半导体

本报讯 记者许子皓报道:近日,有报道称,东芝将在2022财年投资约1000亿日元(约合8.3亿美元)用于扩大石川县工厂功率半导体的制造能力。此外,东芝正在研发将存储容量提高至目前的1.7倍、达到30TB以上的技术。

据了解,东芝2022财年的投资金额比2021财年所预期的690亿日元(约合5.7亿美元)增加了约45%。计划将于2023年春季开始,在生产子公司加贺东芝电子的场地内新建生产厂房,引进新设备,而且现有的生产厂

房内也将增添新的生产线。东芝表示,使用碳化硅或氮化镓的新一代功率半导体的产能将增加。东芝的目标是将功率半导体的整体产能提升2.5倍。东芝认为,功率半导体有助于抑制电力损耗,随着双碳目标的逐渐落地,需求正在增加。

赛迪顾问集成电路中心负责人滕冉向《中国电子报》记者表示,功率半导体是电能转换与电路控制的核心,下游应用广泛。近年来,随着社会经济快速发展和技术工艺的不断进步,功率半导体的应用领域已从传统的工业控

3月22日晚11点,英伟达CEO黄仁勋在GTC大会上又演讲了,演讲地点从自家厨房搬到了公司。

此次演讲,老黄将关注的重点聚焦在AI上。100分钟的演讲共提及161次AI,从英伟达当前支持的AI应用,到更支持AI技术实现的处理

器,再到英伟达提供的AI平台Omniverse。看来老黄这次是打算跟AI死磕了。

英伟达又放AI大招

本报记者 姬晓婷

天气预报 AI 模型 提前一周预测灾难性天气

“传统的数值模拟需要一年的时间,而现在只需要几分钟。”黄仁勋介绍称,英伟达与包括加州理工学院、伯克利实验室在内的多家科研机构合作开发的FourCastNet天气预报AI模型,将能够预测飓风、极端降水等天气事件。黄仁勋称,FourCastNet由傅里叶神经算子提供动力支持,基于10TB的地球系统数据进行训练。依托这些数据,以及NVIDIA Modulus和Omniverse,可实现提前一周预测灾难性极端降水的精确路线。

不仅是在极端天气愈加频繁的情况下发挥作用,英伟达的产品也因疫情而愈加普遍化的在线办公更加智能化。配合在线会议的发展,黄仁勋在演讲中正式发布NVIDIA Riva。这是一种先进且基于深度学习的端到端语音AI,可以自定义调整优化,已经过预训练,客户可以使用定制数据进行优化,使其学习特定话术,以应对不同行业、国家和地区的需求。

另一种为应对在线办公而生的SDK(Software Development Kit,软件开发工具包)Maxine,也在黄仁勋此次视频演讲中呈现。这是一个AI模型工具包,目前已拥有30个模型,可以帮助用户在参与线上会议的时候与所有人保持眼神交流,即便是正在读稿也不会被发现,还能实现语言之间的实时翻译。

“这是全球AI计算基础架构引擎的巨大飞跃,下面隆重推出NVIDIA H100。”在演讲中,黄仁勋再次推出新产品。H100采用TSMC 4N工艺,具有800亿个晶体管,是首款支持PCIe 5.0标准的GPU,也是首款采用HBM3标准的GPU,单个H100可支持40Tb/s的IO带宽。从另一个角度来说,20块H100 GPU便可承托相当于全球互联网的流量。Hopper架构相较于前一代Ampere架构实现了巨大飞跃,其算力达到4PetaFLOPS的FP8,2PetaFLOPS的FP16,1PetaFLOPS的TF32,60TeraFLOPS的FP64和FP32。H100采用风冷和液冷设计,据黄仁勋介绍,这是首个实现性能扩展至700瓦的GPU。在AI处理方面,Hopper H100 FP8的4PetaFLOPS算力是Ampere A100 FP16的6倍。

“搭积木”技术 建成 AI 工厂

“这是全球AI计算基础架构引擎的巨大飞跃,下面隆重推出NVIDIA H100。”在演讲中,黄仁勋再次推出新产品。H100采用TSMC 4N工艺,具有800亿个晶体管,是首款支持PCIe 5.0标准的GPU,也是首款采用HBM3标准的GPU,单个H100可支持40Tb/s的IO带宽。从另一个角度来说,20块H100 GPU便可承托相当于全球互联网的流量。Hopper架构相较于前一代Ampere架构实现了巨大飞跃,其算力达到4PetaFLOPS的FP8,2PetaFLOPS的FP16,1PetaFLOPS的TF32,60TeraFLOPS的FP64和FP32。H100采用风冷和液冷设计,据黄仁勋介绍,这是首个实现性能扩展至700瓦的GPU。在AI处理方面,Hopper H100 FP8的4PetaFLOPS算力是Ampere A100 FP16的6倍。

不仅注重速度和算力,H100也注重数据使用的安全性。

“通常,敏感数据处于静态和在网络中传输时会进行加密,但在使用期间却不受保护。”黄仁勋假设了一个场景,若一家公司具有价值数百万美元的AI模型,而在使用期间不受保护,则该公司将面临巨大的数据风险。他声称,Hopper机密计算能够保护正在使用的数据和应用,能够保护所有者的AI模型和算法的机密性以及完整性。此外,软件开发者和服务提供商可在共享或远程基础架构上分发和部署宝贵的专有AI模型,在保护其知识产权的同时扩展业务模式。

黄仁勋隆重发布的全新AI计算系统DGX H100展现出英伟达像搭积木一样拓展处理器性能的技术。借助NVLink连接,DGX使8块H100成为了一个巨型GPU:拥有6400亿个晶体管,具备32PetaFLOPS的AI性能,具有640GB HBM3显存以及24TB/s的显存带宽。

仅仅连接GPU还不过瘾,英伟达“搭积木”的技术可以再将8块GPU组成的DGX进行连接。黄仁勋推出NVIDIA NVLink Switch系统,借助NVLink Switch系统,计算系统可扩展

为一个巨大的拥有32个节点、256个GPU的DGX POD,HBM3显存高达20.5TB,显存带宽高达768TB/s。每个DGX都可借助4端口光学收发器连接到NVLink Switch,每个端口都有8个100G-PAM4通道,每秒能够传输100GB数据,32个NVLink收发器可连接到1个机架单元的NVLinkSwitch系统,以此实现超强的拓展性。

黄仁勋称,英伟达正在建造EOS——英伟达打造的首个Hopper AI工厂,搭载18个DGX POD、576台DGX、4608个H100 GPU。在传统的科学计算领域,EOS的速度是275Pet-aFLOPS,比A100驱动的美国速度最快的科学计算机Summit还快1.4倍。在AI方面,EOS的AI处理速度是18.4Exa-FLOPS,比全球最大的超级计算机——日本的Fugaku快4倍。

从H100到使用8块H100构成的AI计算系统DGX H100,再到使用256个GPU的DGX POD,以至于HopperAI工厂,英伟达像搭积木一样,构建起一套辅助AI计算的硬件系统。

与英特尔打擂台的 Grace 有望明年供货

在去年的GTC大会上,英伟达推出了首颗数据中心CPU——Grace。按照英伟达的介绍,这是一颗高度专用型处理器,主要面向大型数据密集型HPC和AI应用。与英特尔CPU坚守的x86架构不同,Grace另起炉灶采用ARM架构。黄仁勋声称,服务器用上这款CPU后,AI性能将超过x86架构CPU的10倍。

此次GTC大会,黄仁勋称Grace进展飞速,有望明年供货。不止于此,他们将“搭积木”技术继续应用在了Grace技术上。通过Grace与Hopper连接,英伟达打造了单一超级芯片模组Grace-Hopper。黄仁勋称Grace-Hopper的关键驱动技术之一是内存一致性芯片之间的NVLink互连,每个链路的速度达900GB/s。Grace CPU也可以是由两个通过芯片之间的NVLink连接、保证一致性的CPU芯片组成的超级芯片,可拥有144个CPU核心,内存带宽高达1TB/s。

接着,老黄给出了Grace和Hopper能够打造的不同排列组合方案:2个Grace CPU组成的超级芯片、1个Grace加1个Hopper组成的超级芯片、1个Grace加2个Hopper的超级芯片、搭载2个Grace和2个Hopper的系统、2个Grace加4个Hopper组成的系统、2个Grace加8个Hopper组成的系统等。

“老黄”与“小黄”的对话 透露何种玄机

老黄的这次发布会,再次请出了英伟达仿照自己的形象设计的虚拟人——Toy Jensen。而这次,虚拟人Toy Jensen出现的主要目的,是展示英伟达用于构建虚拟形象或数字人框架的Omniverse Avatar。

Toy Jensen在完成了一轮百科功能展示之后,又兴致勃勃地站在老黄对面展示起了自己的出生地——Omniverse Avatar。这是一个基于Omniverse平台构建的框架,用户可以快速构建和部署虚拟形象。“小黄”Toy Jensen的声音、面部均由英伟达的系列工具提供。“小黄”的声音由Riva的文本转语音RADTTs合成,Omniverse的动画图形可定义并控制其动作,Omniverse Audio2Face可驱动其面部动画。NVIDIA的开源材质定义语言(MDL)可增加触感,使“小黄”的衣服看起来更有合成皮革的视觉感受,而不仅仅是塑料。最终,“小黄”的形象通过RTX渲染器能以实时高保真的程度呈现。得益于Riva中的最新对话式AI技术和Megatron 530B NLP模型,“小黄”可以与真人进行对话。不仅如此,归功于一款使用Omniverse Avatar构建的应用Tokkio,“小黄”还能连接到更多类型的数据,它将客户服务AI引入零售店快餐餐厅,甚至网络。