

# 新型数据存储仍待突破“最后一公里”

■本报记者 张双虎

信息的“洪流”可能很快就会超过硬盘的承载能力。

大数据正像发动机一样赋能千行百业，随之而来的是数据体量的爆炸式增长。然而，当前用于存储数据的设备，比如光、电、磁介质和闪存，使用寿命通常不到20年，并且还需要消耗大量能量来维护。

尽管科学家们始终在探索以高密度和无能耗的方式来存储信息的新方法，比如DNA或其他分子聚合物存储技术，但其相对成本高、读写速度慢等瓶颈问题仍未突破。近期，几项关于存储技术的新进展给数据存储带来了一线曙光。

## 新的光学分子存储

近日，哈佛大学化学与化学生物系教授George Whitesides团队及其合作者，开发了一种全新且原理简单的光学分子存储方法。他们使用荧光染料分子编码二进制信息，成功将各类文字、图像信息写入并读取出来，其读、写速度分别达到每秒128比特和469比特，甚至可以读写视频或其他信息。相关研究已在《ACS核心科学》上发表。

光学分子存储技术不需要对分子内部的结构或序列进行复杂编码，且在理论上具有较高的存储密度、无需能耗维护。然而，长期以来读、写速度慢，读取次数有限等问题也制约着其进步发展。

为解决此类问题，Whitesides等人使用荧光分子将目标信息编码成二进制并存储，并使用荧光显微镜来读取信息。使用这种方法，他们成功在一块小至7.2平方毫米的基板上存储了约14KB的信息，空间信息密度达到每平方米271.5字节。此外，通过重复读取信息1000次以上，研究团队发现荧光信号强度的损失较少(低于20%)，且每次都能成功读取。

“Whitesides团队的存储方法原理简单、概念新颖，在信息存储方面具有一定的意义。”华中科技大学化学与化工学院教授吉晓帆告诉《中国科学报》。

中国科学院自动化研究所模式识别国家重点实验室研究员杨戈同样认为，“这是一个创新的方法”。

“它实际上是基于化学荧光分子的存储技术。”杨戈对《中国科学报》说，“但它用喷墨打印来实现，存储密度、读取速度可能会受到影响。此外，通过荧光标记的方式，会面临荧光分子的发光衰减问题，从而可能限制读取次数。”

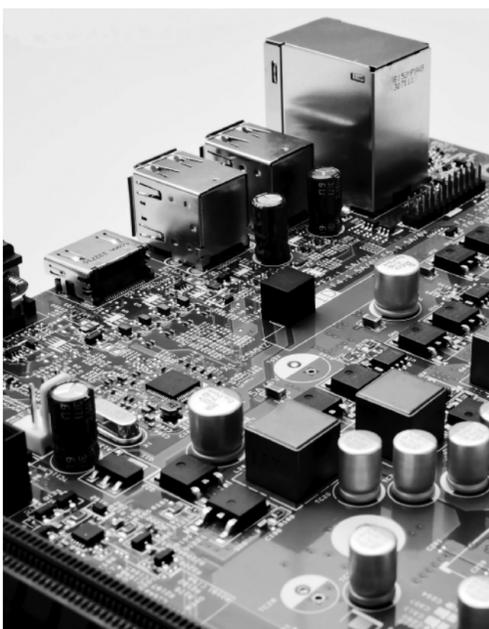
## 按需开发存储材料

“信息存储手段多种多样，不同存储手段

信息存储手段多种多样，不同存储手段有不同的优点，可以应用于不同场景。

人们可以根据自己的需求，如存储信息的内容、安全性等，选择存储手段，开发不同的存储材料。

图片来源: unsplash



有不同的优点，可以应用于不同场景。”北京化工大学材料科学与工程学院教授郭金宝告诉《中国科学报》，“人们可以根据自己的需求，如存储信息的内容、安全性等，选择存储手段，开发不同的存储材料。”

近日，郭金宝团队在《先进功能材料》发表文章，使用具有可重写荧光图案和可重构3D形状的液晶弹性体材料来实现多重信息存储与加密。他们利用材料特性产生的两种信息存储方式，进一步拓展了存储信息的数据量；另一方面，该体系在读取信息时需要两步不同的解密方式，使存储信息的安全性大大提高。

“需要指出的是，我们这个工作属于基础研究，旨在拓展液晶弹性体在信息存储方面的应用，距离实际应用还存在较大差距，仍需进一步探索。”郭金宝说。

而吉晓帆团队在线发表于《先进材料》的成果，则展示了利用聚集诱导发光(AIE)超分子水凝胶实现信息大量存储的可能。

“AIE超分子水凝胶是以信息码技术为基础，结合AIE分子、超分子凝胶的优良性能，集光、通信和网络技术为一体的新型存

术，除了存取密度、速度、使用寿命和安全性外，还要考虑其稳定性、方便性和成本。”

## 产学研合作加速产业化

“现在确实面临一个问题。”杨戈说，“我们生活在一个信息化的社会，对数据的需求越来越多，产生的数据量越来越大。比如，腾讯、阿里、字节跳动等信息技术公司，在能源价格低廉的偏远省区建了很多数据中心，但能源消耗仍然是个大问题。而且，现在的主流存储方式都有其物理上的限制，人们迫切需要寻找新的存储技术。”

杨戈介绍说，DNA存储技术用核酸遗传密码的4种核苷酸(即腺嘌呤、胸腺嘧啶、鸟嘌呤和胞嘧啶)将信息编码在DNA中。其基本原理是一次写入、多次读出，是一种长期存储的技术。目前DNA存储技术虽然有很多不同的实现方式，但基本的技术路线是一致的。

借助现代生命科学领域的成熟技术，比如合成和测序等，DNA存储技术得到快速发展。美国微软、因美纳等数据存储和生物技术公司还成立了DNA存储技术联盟，共同推动该技术的发展。

杨戈认为，数据存储技术有很多种，单就DNA存储技术来说也有多种实现方式，将来谁做得比较好，谁可能就会闯出一条路来。美国加入DNA存储技术联盟的公司和大学已经有三四十家，“从长远看，这有可能又会成为一项关键技术”。

“DNA作为数据存储材料具有存储量大、存储密度高、存储周期长等优点。”郭金宝说，“然而，该技术在当前面临的挑战还很大，如合成成本高、读取过程复杂等，并且其相关研究主要停留在实验室层面，距离真正的应用还有很长的路要走。”

截至目前，我国在DNA存储上已经形成了一批队伍，中国科学院国家纳米科学中心、天津大学、南方科技大学等单位都有课题组在做相关研究。

“我们目前在基础研究方面已经有了不错的积累，在纳米技术方面也走在国际前列。”杨戈说，“但在实验室里做出一个原理性的东西，发几篇论文，与投入实际应用还有很长一段距离，其中有很多复杂的工程问题需要解决。下一步，希望能够成立技术联盟，通过研究机构和企业的合作，加速技术产业化。”

相关论文信息：  
<https://doi.org/10.1021/acscentsci.1c00728>  
<https://doi.org/10.1002/adma.202105418>  
<https://doi.org/10.1002/adfm.202107145>

# 从大奖赛到计算应用交流平台：交叉计算人才在这里成长

■本报记者 赵广立

作为中国计算领域的行业盛事，以“纵横驰骋‘码’上争锋”为主题的第二届“先导杯”计算应用大奖赛暨颁奖典礼(以下简称“先导杯”)，于近日在北京收官。

毕业于北京大学的“香蕉牛奶”战队，今年是第二次参加“先导杯”。在去年的首届“先导杯”中，他们(首届战队名为“榴莲牛奶”)就是最大赢家，一口气揽获基础算法赛道三个赛题的特等奖；今年，他们将四个赛道的两个一等奖和两个二等奖收入囊中，再度成为本届“先导杯”最多奖金获得者。

其实他们并不在意奖金多少。“我们觉得‘先导杯’是一个很好的平台，在这里我们可以学习硬件怎么搭，看看其他参赛者有什么好想法，也认识了很多来自不同领域的朋友。”

相比去年参赛，他们觉得今年拿奖相对轻松些。去年参赛沉淀的方法今年还很有用，这让他们事半功倍。“我们今年能拿奖，确实沾了去年的光。”“香蕉牛奶”战队成员告诉《中国科学报》，“我们准备整理一下方法和工具链，希望开源出来，让更多的人可以使用并丰富它们。”

可以看到，两岁的“先导杯”已经从大奖赛变成了一个计算应用技术、人才交流平台。

## “PK平台”的催化剂效应

“先导杯”组委会秘书长李斌很喜欢“催化剂”这个提法。

“大赛就像提供了一种催化剂。我和很多选手沟通，发现在赛题的设置上，有些院校机构也有这样的研究方向，也要做类似的人才培养等工作。大赛的催化作用在于有奖金——这还不是主要因素，重要的是大赛提供了一个PK的平台，这样大家的动力就更足。”李斌说，“当然，有竞争，就会有交流。”

在竞争、交流的氛围下，两届“先导杯”涌现了一批有关计算应用方法、算法优化等成果，对于计算软件应用乃至生态培育都起到了积极的促进作用。

其实从工作量来说，对赛题算法的功能实现，只是完成了“软件开发”的10%~20%，而要实现软件的产品化，还有大量的工程化、用户体验优化等问题。

“做一个软件，即使算法特别好，但如果没有人用，它就没有生命力。”李斌说，这是当前一个大问题。

在他看来，有两种渠道解决这一问题：一是成功的商业化开发；二是开源，让更多人使用、参与开发。所以，当他听到“香蕉牛奶”战队要开源他们的方法及工具链时，“非常欢迎”。

作为主办方，李斌也在考虑：“‘先导杯’目前还只是一个大赛平台，后续我们可能会赋予这个平台更多的功能，包括做产业孵化之类的合作。”他觉得，这可能会让大赛“发生质的变化”。

## 赛题设置优化显成效

在首届“先导杯”颁奖典礼上，中国工程院院士李国杰看到“我国有上千所大学开设了计算机专业，但是参加这次竞赛的团队大都来自重点大学”，提出“并行计算不是‘阳春白雪’，而是计算机从业者需要掌握的基本功……希望更多的研究生从事并行计算的科研和技术攻关”。

大赛的举办正起到引起更多关注的作用。据李斌介绍，第二届“先导杯”的关注度和参与度相比首届都有了明显变化：“今年无论是参赛选手、参赛队伍还是涉及行业领域的数量相比首届都有了两倍以上提升。”

“众人拾柴火焰高。”李斌分析，这背后除了群策群力之外，至少还有两方面原因：一是首届“先导杯”开了一个好头，积累了口碑；二是今年办赛的合作组织方式进一步改进，对赛题的设置做了优化。

在赛题设置上，“先导杯”既覆盖了传统的科学与工程计算，也包含诸如量子计算等新应用，同时覆盖基础算法和上层应用。

“指导老师想尽办法，让赛题一方面更加有趣，另一方面更加贴近生产、生活中的实用问题。”李斌举例说，今年的人工智能应用赛题选取的是非常有实用价值的“智能交通检测”，并提供百度飞桨平台作为开发环境；而量子计算赛题的设置，目的是让大家知道，用经典计算机也可以研究许多量子计算前沿问题。同时，大赛还延续了有着命题作文特色的开放应用赛道。

大赛中，“Seigato”战队结合一站式人工智能开发平台，分别在自动驾驶、智能导航和

城市道路环境变化监测方面提供了解决思路及方案，对于实时地图导航场景下，不同种类的交通检测的实时提醒、规范车驾行为、保障用户出行等方面有重要借鉴意义。而来自中国科学院软件研究所的“ISCAS-SC”战队，其提交的应用项目，通过精准的对流数值模拟算法，对水汽运动进行特殊传输模拟，可为数值天气预报以及雾霾预警提供算法及应用支持。

“大赛对于解决我国一些重大需求问题也有促进作用，因为很多赛题都带有一定的挑战性和通用性。”北京科技大学计算机科学与技术系教授胡长军是“开放应用”赛道的评委之一。他举例说：“比如赛题中的稀疏矩阵求解、特征值问题等，是重大工程应用中的核心和基础问题。这些问题解决好，有利于促进许多战略工程的进展。”

## 大赛成果希望有后续效果

“先导杯”让中国科学院国家天文台副研究员王乔感触颇深的两点是，第一，参赛选手的合作意识；第二，大赛上的一些成果甚至技术细节都有借鉴价值。

“我的专业偏向于应用领域研究，很难培养专门的计算人才，如何把计算引入进来变得非常重要。而大赛就是一个较好的方式，能够把人才和技术积累下来，为我们提供借鉴和帮助。”王乔说。

在中国科学院计算机网络信息中心总工程师迟学斌看来，虽然大赛涌现的成果离实际应用还有一段距离，但未来这些方法和模拟如能与职能部门数据结合，比如做药物设计模拟的与制药公司结合，就能发挥巨大的现实作用。

迟学斌是本届“先导杯”的专家评委会主席。他看到，首次参赛的青岛大学战队，就用计算机做了雅鲁藏布江地区的水文模拟，“这

类工作对于农业生产、气候气象等会起到重要作用”。

“特别是现在我们拥有了强大的计算能力，以前不敢想象的模拟也可以做了。”迟学斌说，“先导杯”的目的之一在于把大规模计算机的规模和性能展示出来。“并不是要求大赛成果一定要落地应用，而是通过这些工作做一些积累，相信后续会在一些应用领域产生效果。”

## 培养交叉复合型人才

谈及大赛的人才培养作用，迟学斌看得比较长远。

“一次比赛不一定一下子就能培养多少人，只是通过这种形式告诉大家，并行计算应用这件事情，应该去注重；而且通过大赛的形式，让大家更清楚地知道，科研中需要这种手段。”

迟学斌同时认为，“先导杯”强调了复合型人才培养的作用和意义。“通过五六个月的培训和学习，参赛选手可能会逐渐意识到，用计算手段做科研，需要付出多少努力，因为除了物理、材料等本学科的工作之外，还要做一些计算机方面的工作。”

“我们的计算人才还是很缺乏的。”胡长军说，大赛除了对学生实实在在的编程能力进行了验证，还考验了他们的敬业精神、吃苦耐劳精神。但这还不够，通过参赛，让参赛选手尤其是学生选手，踏踏实实、认真真学学怎么做科研，如何解决这些问题，这也是很重要的。

对于大赛对人才交叉综合能力的培养，王乔从计算应用领域代表的角度，提出了他的一点希望，“每次比赛都会积累很多宝贵的知识和经验，我们应该把它提炼出来，放进‘社区’，让其他院校机构的学生、科研人员触手可及，甚至参与进来，这样有助于计算与应用的融合和交叉人才培养，产生更广泛的影响。”

本报(记者赵广立)在近日举行的2021 CCF全国高性能计算学术年会上，三场介绍超算大规模应用的特邀报告引起了与会者的广泛关注。这三项中国超算应用目前已入围国际上高性能计算应用领域的最高学术奖项——“戈登贝尔奖”，并将于11月中旬举行的全球超级计算大会(SC21)上冲击该奖项。

入围“戈登贝尔奖”的三项超算应用分别为：超大规模量子随机电路实时模拟的应用(报告人单位：国家并行计算机工程技术研究中心)、千万核可扩展第一性原理拉曼光谱模拟(报告人单位：中国科学院计算技术研究所)、多架构大规模并行保辛结构电磁全物理等离子体模拟(报告人单位：中国科学技术大学核科学技术学院)。

会上，国家并行计算机工程技术研究中心研究员、神威系列智能计算机常务副总设计师刘鑫对“硬件—软件—应用”协同构建的新一代神威国产超级计算机系统进行了介绍，并报告了运行其上的“超大规模量子随机电路实时模拟”。她透露，该应用可在60小时内完成100量子比特40层深度的量子随机电路的单振幅模拟，304秒内完成谷歌“悬铃木”当年“量子霸权”问题的模拟(谷歌当年宣称，使用美国超级计算机Summit要得到类似的结果需要一万年)。

中国科学院计算技术研究所副研究员商红慧介绍，该团队重新设计了全电子精度的第一性原理计算理论算法，提高了生物蛋白质体系模拟算法的精确度和适用广泛度，使得通过量子力学方法直接模拟得到蛋白质结构信息成为可能，这将为分析药物的动力学过程开辟一条新途径，有望使药物的研发更快速、更精确。展望未来，她认为，学术界和产业界都已意识到量子力学方法在生物蛋白质体系模拟中的重要性。随着算力的不断提升，该方法的应用越来越广泛，“越来越多化学家不用去实验室，直接奔赴机房就可以了”。

中国科学技术大学核科学技术学院副研究员肖建元在特邀报告中介绍，高精度、高分辨率的等离子体模拟对于建设用于核聚变研究的工程意义重大，而该模拟需要算法具有高可扩展并行性、高单节点性能，尤其是须保证算法稳定可靠、长时间守恒。他们团队用基于scheme语言开发的适用于“主—从架构”的领域专用语言PSCMC，实现的大规模并行保辛结构PIC程序SymPIC，通过在新一代神威平台上的优化，提升了在多种异构架构超算平台上的可移植性与调优，并保证了长时间大规模计算的可靠性。

一位熟悉“戈登贝尔奖”的超算专家告诉《中国科学报》，这三项应用的软件水平高、方法有创新且所使用的硬件设备规模足够大，因此对它们冲击“戈登贝尔奖”保持乐观，认为“应该有机会”。

设立于1987年，由美国计算机协会与美国电气电子工程师协会联合颁发的“戈登贝尔奖”，是国际高性能计算领域的最高学术奖项，有“高性能计算领域的诺贝尔奖”之称。该奖主要颁发给高性能应用领域杰出贡献者，目的是表彰他们在利用并行计算实现科学突破方面所取得的成就。在2016年之前，美国、日本两大超算强国垄断该奖项长达近30年。

中国超算应用于2016年实现了对“戈登贝尔奖”的“破冰”，此后又先后于2017年、2020年摘得这一桂冠。其中，2020年的“戈登贝尔奖”颁给了一支主要由中国科学家构成的中美研究团队(8位获奖者中有7张中国面孔)。

## 速递

### 中关村5G创新应用大赛举行

本报讯 日前，2021年中关村5G创新应用大赛总决赛暨第四届“绽放杯”5G应用征集大赛京津冀区域赛在京举行。“后疫情时代的5G全连接智能物流交通建设和物流中心应用”“基于5G+AI的交视视频融合感知及数字孪生公路系统”两个项目获得一等奖。

据悉，中关村5G创新应用大赛此前已连续举办三届。本届中关村5G大赛联合天津、河北两地产业主管部门共同主办，聚焦挖掘关键元器件、虚拟专网及垂直行业的创新项目。在大赛推动下，行业单位积极探索5G在不同场景的融合应用，涌现出一大批智慧城市、智能制造、智慧医疗等领域优质项目。

大赛共吸引京津冀申报项目近90项，项目投资总额近40亿元。其中，北京提供整体解决方案、津冀复制应用落地项目超过15项。

(郝金武)

### 新技术提高机器人团队应急响应速度

本报讯 机器人团队可以帮助用户更快、更高效地完成大量任务，并在危险操作期间使人类特工免受伤害。近日，美国布法罗大学的研究人员在《Robotics and Autonomous Systems》期刊上发表的一篇文章，介绍了一种可以提高机器人团队在灾难响应任务中表现的技术。这种技术旨在将任务分配给团队中的不同机器人，以便它们能够最有效地完成任务。

目前现有的多机器人任务分配方法很少能够处理对于时间有严格要求的并发任务，更难以应付任务期间可能出现的意外。与之相比，研究人员开发的新技术执行时间更短，可以在几百毫秒内作出任务分配决策。同时，新技术降低了团队中机器人对通信网络的依赖性，从而减少机器人之间同步决策的时间。

(袁一雪)

相关论文信息：  
<https://doi.org/10.1016/j.robot.2021.103905>