intel®

Novel Compute Capabilities for Mass Spectrometry Solutions

Equipment providers can differentiate their mass spectrometry solutions for the lab and life sciences segment by supporting capabilities and usage models based on Intel architecture, including increased test volume, enhanced reliability and federated learning.

The market outlook for mass spectrometer equipment continues to be vigorous, with a projected CAGR of 8.2% through 2031, reaching $13.5 billion that year.[1] Growing research and development investments by the pharmaceutical and biotechnology industries is a major contributor to that trend, with ever-increasing sophistication in development and discovery methods toward novel drugs, biopharmaceuticals and vaccines. In fact, this segment is responsible for 18.9% of the global expenditure for research and development overall.[1]

Likewise, advances in genetics (including work on gene therapies), personalized medicine and increased uptake by the food and beverage industry contribute to the ongoing growth in demand for mass spectrometers. Some common applications for mass spectrometry are shown in Figure 1. Other key areas where it is broadly used include industrial chemistry and environmental testing.


Figure 1. Common applications for mass spectrometry.

Computer systems are embedded in spectrometry instruments themselves, supported by edge computers that provide analysis and control functions, as well as systems for large-scale analytics and data collaboration. Equipment makers can differentiate their offerings and offer multi-level solutions with features and capabilities based on Intel architecture. Holistic development and services improve research workflows for lab and life sciences customers, for improved efficiency and outcomes.

Research Context for Mass Spectrometry

The core functional sequence of mass spectrometry — ionizing molecule mixtures, separating ions with electromagnets and determining the mass of molecules — is applicable to hundreds of "omics" applications. Key research fields in this category (in terms of the use of mass spectrometry) are illustrated in Figure 2. Within the larger sphere of lab and life sciences, proteomics is a primary usage area for mass spectrometry, as is metabolomics.
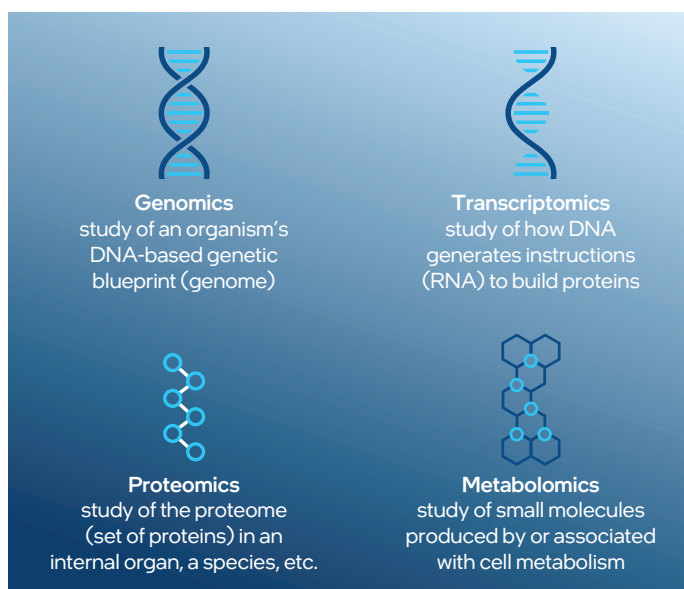
Figure 2. Example fields of research enabled by mass spectrometry.

As the large-scale study of all proteins in a biological system, proteomics can be used, for example, to quantify the chemical effects of a drug on the proteome (set of proteins). The impact of mass spectrometry can be demonstrated in the comparison of the proteome of a cellular sample that has been treated using a given drug with the proteome of a non-drug-treated sample. The results may reveal elevated or suppressed levels of individual proteins, as well as chemical modifications within the proteins themselves. Quantifying these translational effects is critical to advancing drug discovery, helping characterize the behavior of specific compounds on biological systems under certain conditions, such as a disease state. Mass spectrometers are used throughout this process and are a critical piece of this form of scientific discovery.

Opportunities to Improve Research Outcomes

Growing research costs and complexity make it increasingly critical to get the maximum value possible from investments in equipment and personnel. The computing technology involved in mass spectrometry workflows offers opportunities for both equipment makers and their customers to improve data handling for efficiency and flexibility. The key opportunities shown in Figure 3 are each discussed in greater detail in the remainder of this section.
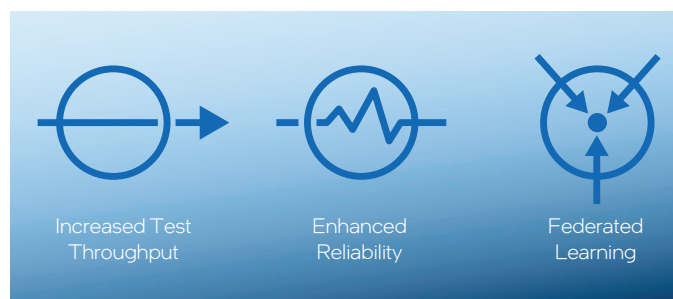
Figure 3. Improvement opportunities for mass spectrometry using Intel architecture.

Increased Test Throughput

The high capital costs associated with mass spectrometry equipment make it vital to get the most value possible from each instrument. Labs cannot easily purchase additional mass spectrometers, making them a precious resource that must typically be shared among teams. For example, in proteomics, the typical usage is to load a cartridge full of samples into an auto-sampler and to run the instrument for an average of approximately 90 minutes. The instrument itself is therefore dedicated to that job for an extended period, creating a potential bottleneck that could impact how quickly the study can be completed, while other studies must queue up waiting their turn. Measures to increase throughput and reduce that bottleneck are therefore an important opportunity.

In practice, a mass spectrometer is typically deployed in conjunction with an external edge computer, as illustrated in Figure 4. Cell or tissue samples are prepared for proteomic workflows by disrupting cellular membranes to separate proteins, which can then be isolated from lipids and metabolites. Proteins are then digested into a peptide digest mixture using enzymes. The data acquisition stage begins by separating specific peptides from the mixture using liquid chromatography and performing mass spectrometry on each, in real time. After performing a survey scan, the mass spectrometer sends the data set to the edge computer. The device analyzes the data and provides direction to the spectrometer to perform peptide fragmentation. This process separates the peptides into individual amino acids to enable sequencing.
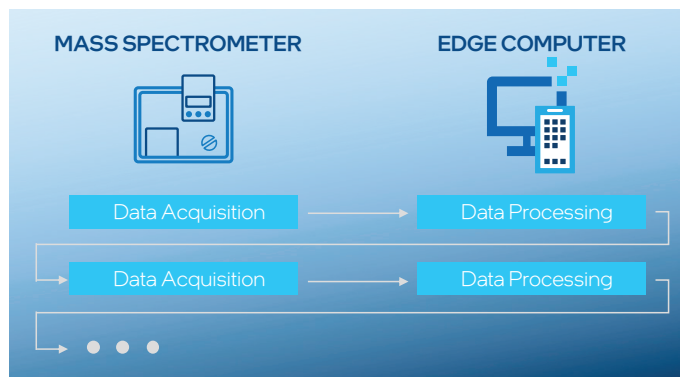
Figure 4. Iterative data acquisition and processing.

Therefore, the data flow is iterative, with the output of the mass spectrometer acting as the input for the edge computer and vice versa, with a cycle time of a few seconds. In this scenario, edge compute is typically the rate-determining step, meaning that the instrument must continually wait for the edge computer to complete its work, preventing it from running at optimal capacity.

Increasing the computational throughput provided by the edge computer allows more experiments to be performed per cycle, as well as decreasing the cycle time so more sampling occurs. These factors help make more efficient use of scarce instrument resources. Because even modest improvements in compute speed can equate to a substantial time savings over thousands of data acquisition and processing cycles, relatively small investments in more robust edge compute can add significant end-customer value for researchers. Holistic solutions developed by instrument makers can benefit from the use of Intel's high-performance architecture-based edge computers to deliver that value, for a potential competitive advantage.

Intel, in partnership with suppliers and manufacturers, provides hardware and software that is tailored to the needs of such implementations. Processors offer high per-core performance and are available with high core counts and features that help optimize these workloads. The high performance of balanced edge computer systems are enhanced with additional hardware ingredients such as Intel® Ethernet intelligent network adapter and accelerators. In addition, solutions draw from Intel's software and solution ecosystem, including open source contributions as well as partnership with proprietary hardware and software makers.

Offline data processing is similarly optimized on large-scale systems, with proteomics data largely still held in on-premises storage and processed by on-premises computer resources. Distributed computing in this area is in its early stages of adoption but has the potential to increase flexibility and agility. The process begins by sequencing the peptides that have been fragmented by the mass spectrometer and reassembling that data into representations of thousands of proteins. The resulting proteomics data can be analyzed using any of the large number of suitable software tools available, many of which are maintained by the open source community.

Deep learning and AI play a significant and growing role in these analytics, which represents another area of value that can be provided by Intel architecture. In particular, AI provides the ability to automate detection of result trends, reducing cost and time requirements for researchers and scientists. Intel® Xeon Scalable processors, which provide flexible performance based on a variety of core counts and feature sets, offer hardware acceleration capabilities for deep learning algorithms, including Intel Deep Learning Boost (Intel DL Boost) and hardware accelerators. Habana Labs, Intel's data center team focused on AI deep learning processor technologies, offers second-generation dedicated silicon for deep learning: the Habana® Gaudi®2 Training and Habana® Greco™ Inference processors.

Enhanced Reliability

Maximizing the uptime of mass spectrometry equipment is vital to getting optimal ROI from this capital resource and advancing critical objectives and projects. Features and capabilities that enhance system reliability therefore offer equipment makers an opportunity for value-added differentiation of their solutions. Embedded computer systems provide the central management point for mass spectrometry equipment, creating some commonalities between these devices and more general-purpose network endpoints. The equipment can be affected by hardware and software issues from the physical layer all the way upward to the application level that need to be addressed by IT specialists.

Intel Active Management Technology (Intel AMT), a component of Intel vPro® technology, is a hardware-based set of technologies that enhance system manageability and troubleshooting. Spectrometry equipment that incorporates Intel Core™ processors with Intel vPro technology supports an out-of-band management channel that remains operational even if the instrument is powered off, the hard drive has failed or the OS is corrupt or missing, as shown in Figure 5. This channel supports basic system management tasks under these failure conditions that help bring the mass spectrometry equipment back online faster than would otherwise be possible.

Out-of-Band Management Channel

IT Management Console

Mass Spectrometer:
- Powered Off
- Failed Hard Drive
- Corrupted OS
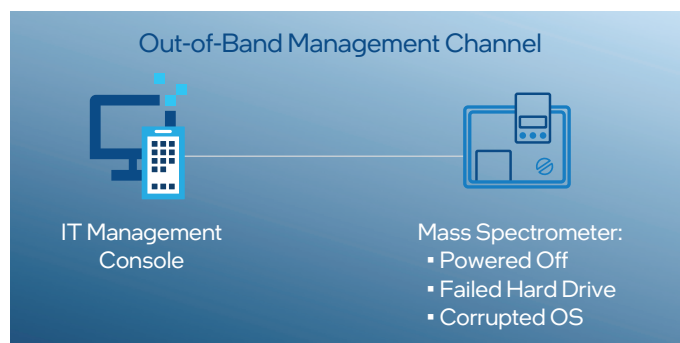
Figure 5. Out-of-band management with Intel Active Management Technology for enhanced reliability.

Intel AMT enables an administrator or technician to access the system from a web-based IT management console separate from the mass spectrometry solution itself, either locally or remotely. That interface enables them to power the instrument on or off, inventory hardware and software components, access system logs and handle various administrative tasks. It can also provide remote keyboard-video-mouse (KVM) control that enables technical staff to interact with the system remotely just as if they were sitting in front of it, regardless of power or OS state. Remote connections can be initiated either by a user or automatically by pre-defined events. They can also occur based on scripting and scheduled connections to a management server for regular maintenance, such as pushing critical patches and other software updates.

With these and other capabilities, out-of-band management can reduce the time and cost requirements to bring systems back online. Equipment makers can offer these additional capabilities in their solutions at little cost, to improve customer outcomes while enabling robust professional services and support offerings.

Federated Learning

Ongoing spectrometry lab research generates large libraries of data. Depending on sample complexity, a typical 90-minute run might generate from 0.5 to 5 GB. That data constitutes valuable intellectual property for the organization that generates it. In addition to the research costs associated with generating those datasets, they can be the basis for developing drugs and other revenue-generating products, the value of which must be protected. In addition, the samples themselves can be difficult to obtain, sometimes requiring meticulous treatment of gifted tissue in the immediate postmortem state, for example. The opportunity to use data for more than one study not only reduces costs but also saves time by making it available for additional research.

At the same time, there are scenarios where making data available on a limited basis can be desirable. Notably, makers of deep learning algorithms often seek access to proteomics data for use in training models. Similarly, collaborative uses of data may arise in the course of solution co-development or other forms of corporate partnering. In the training example, the data itself doesn't need to be retained by the owner of the model. In collaborative partnerships, granular control is often needed over how data is made available outside of its parent organization. Such collaboration addresses myriad goals, from drug discovery, development and label expansion to clinical trial implementation and patient segmenting. It potentially benefits parties as diverse as fellow pharma companies, technology providers and academic researchers.

Federated learning, illustrated in Figure 6, is a sophisticated mechanism for information sharing. This approach allows software to interrogate data in a limited way, without moving it, exposing the data only to trusted sections of application code. Thus, the data itself remains beyond the reach of outside entities. Intel Software Guard Extensions (Intel SGX) is a hardware-based technology built into Intel server processors that enables the creation of isolated, protected regions of memory known as "enclaves."

In this scenario, all data is encrypted unless it is inside a specified enclave, where trusted code can operate on it in its unencrypted form. Untrusted code outside the enclave cannot access the unencrypted data, and the trusted code cannot transfer the data outside the enclave. Thus, the data owner can allow collaborative use without revealing the data itself, therefore protecting intellectual property. Federated learning may also help overcome regulatory restrictions that would otherwise hinder collaboration across geographic and political boundaries.

This model involves transferring algorithms (and their outputs) as opposed to transferring data, which remains protected while in use and never leaves its protected repository. Similar to how a deep learning model can use federated learning to train against data without revealing it to the algorithm owner, analytics can be run on that data to generate collective insights while keeping the data in isolation. Because the protected enclaves are anchored in low-level hardware, they cannot be overcome by software measures, keeping them beyond the reach even of a user or process with root access or a compromised operating system.

Conclusion

Mass spectrometry is a powerful tool that is used across lab and life sciences to enable scientific discovery. It's a resource-intensive research method that involves multi-million-dollar equipment and rarefied technical expertise. Intel architecture and software tools offer the means to increase the efficiency of both inline data processing and offline analytics, offering cost and time-to-market advantages for customers.

High performance in edge computers that support operation of mass spectrometers in near real time can increase the rate of data acquisition, making optimal use of the instruments. Likewise, accelerated offline processing for analytics — including for deep learning — can provide a faster path to discovery and revenue. Hardware-based out-of-band manageability for computers embedded in spectrometers has the potential to enhance reliability of the equipment, helping avoid costly and disruptive outages. Further, AI utilizing the data generated from mass spectrometers can shorten the drug discovery process while increasing the chances of success. These opportunities enable spectrometry equipment makers offer a competitive advantage by accelerating the time to scientific insight and lowering costs.

For additional information, contact your Intel representative at health.lifesciences@intel.com

Lab A — Protected Enclave
Lab B — Protected Enclave
Lab C — Protected Enclave

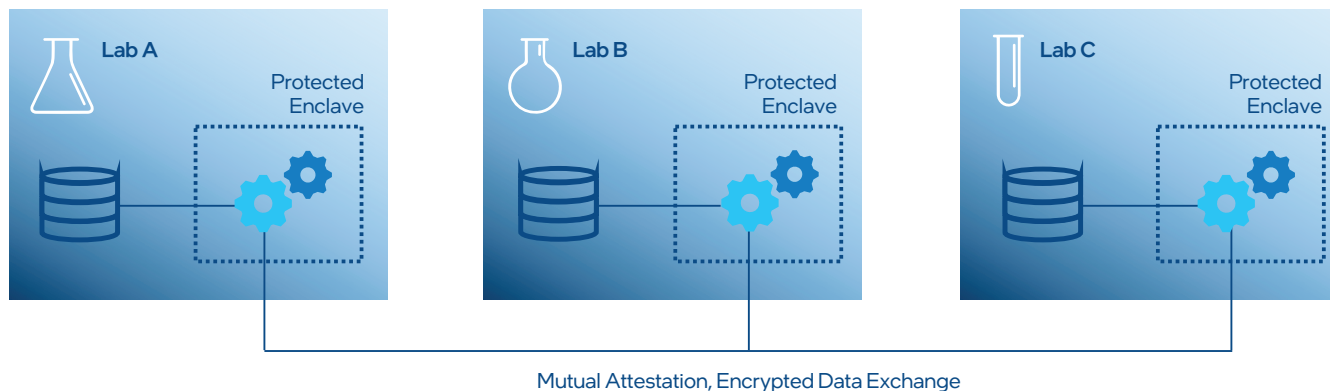Mutual Attestation, Encrypted Data Exchange

Figure 6. Federated learning enables data sharing while protecting intellectual property.

Learn More

Intel Healthcare and Life Sciences Technology Solutions:
https://www.intel.com/content/www/us/en/healthcare-it/healthcare-overview.html

Intel Laboratory Automation:
https://www.intel.com/content/www/us/en/healthcare-it/lab-automation.html

Intel Processors for Embedded Applications:
https://www.intel.com/content/www/us/en/products/details/embedded-processors.html

Intel Active Management Technology:
https://www.intel.com/content/www/us/en/architecture-and-technology/intel-active-management-technology.html

Intel Software Guard Extensions:
https://www.intel.com/content/www/us/en/developer/tools/software-guard-extensions/overview.html

Intel White Paper: "Federated Learning through Revolutionary Technology":
https://www.intel.com/content/www/us/en/financial-services-it/federated-learning-solution.html

Solution provided by:

intel.