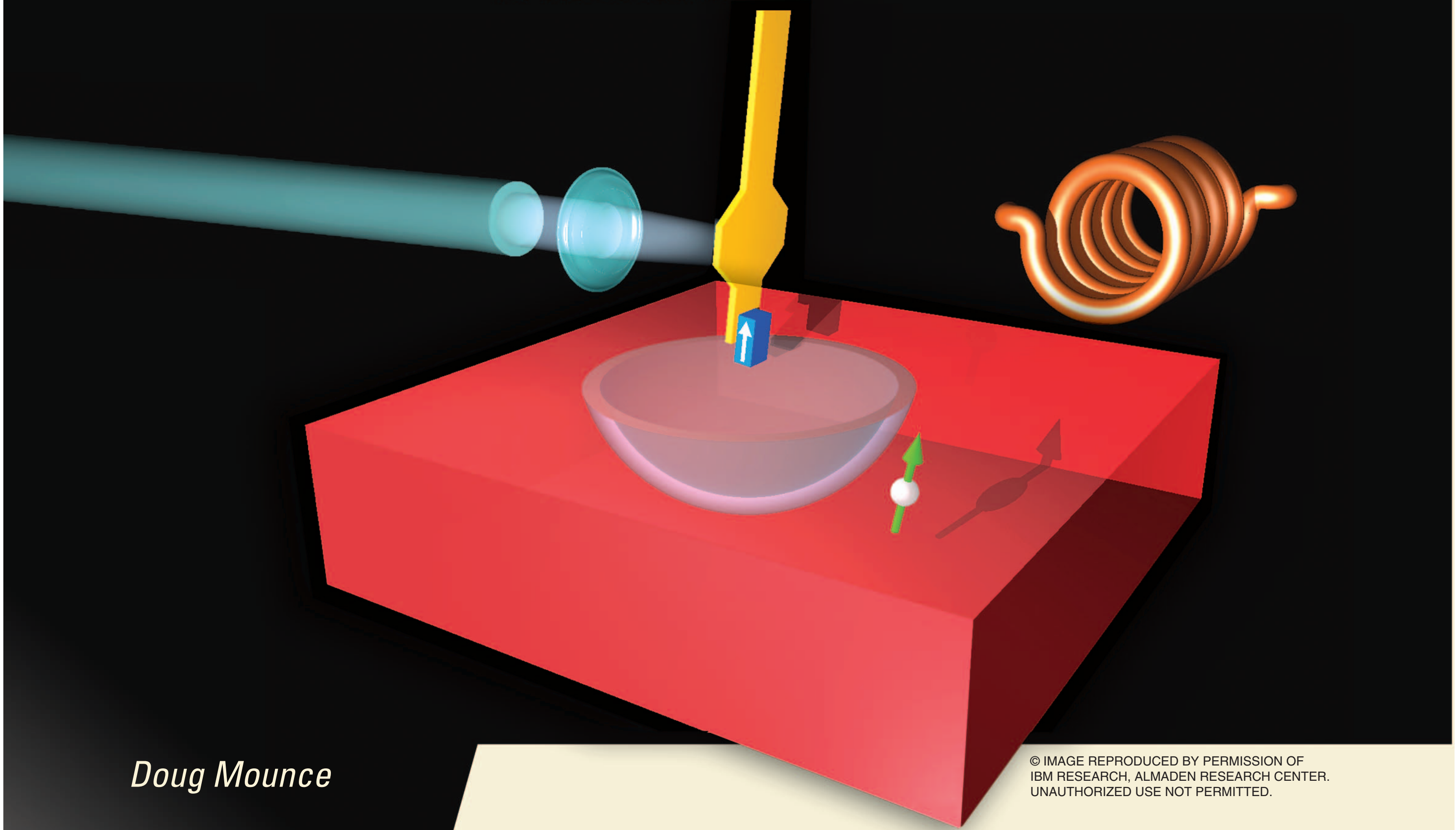
Magnetic Resonance Force Microscopy

This technology's history, unique features, and the software and hardware necessary for its success



© IMAGE REPRODUCED BY PERMISSION OF IBM RESEARCH, ALMADEN RESEARCH CENTER. UNAUTHORIZED USE NOT PERMITTED.

Doug Mounce

Dan Rugar at IBM used magnetic resonance force microscope (MRFM) technology in the summer of 2004 to detect the signal from a single electron spin [1]. This marked a turning point for microscopy since John Sidles invented the MRFM method in the early 1990s [2]. MRFM fills one of the most fundamental gaps in the tools we have for determining the structure of systems and materials with nuclear detection and mapping coordinates at the atomic scale.

"Magnetic resonance force microscopy offers a revolutionary new capability for mapping the composition and structure of molecules or nanostructures with atomic resolution," said Army Research Office (ARO) Program Officers in their Multi-University Research Initiative (MURI) BAA, "even modest improvements will have broad commercial and military utility and will significantly impact advanced semiconductor device research, nanoscience, single-molecule analytical chemistry, biotechnology, and infectious disease research."

This article discusses some history of this microscopy, the unique features of the microscope, and the software and hardware necessary for its success.

1094-6969/05/$20.00©2005IEEE

History

John Sidles, a physicist working in medicine, was inspired during a morning clinical conference to consider what currently available tools could conceivably image those unknown regions of molecular biology where many common diseases elude our detection and understanding. In 1991, the programs intended to move up the scale from genetic information were X-ray crystallography and nuclear magnetic resonance (NMR). Electron microscopes were widely employed, but their high-energy beams disrupted the fragile molecules at the scale of interest. X-ray crystallography is a technique for preserving structure in that environment, but not all proteins crystallize and not all of those yield well-diffracted images. NMR has sophisticated and complex techniques for manipulating the gyroscopic spin behavior but doesn't scale to large molecules or combinations of them. Atomic force microscopy (AFM) is limited to surface imaging and by electrostatic forces. Magnetic resonance imaging (MRI), while advancing from the large-scale down to micrometer-sized particles with techniques like fluorescent spectroscopy and spin-labeling, still relies on the contrast from huge numbers of spins. Dr. Sidles' 1991 publication described how a mechanical lever like those in an AFM could be used for detecting individual spin resonances with NMR techniques in a nanoscale MRI device.

In Richard Feynman's celebrated 1959 address during the American Physical Society (APS) meeting at Caltech (often credited as inspirational for the nanotechnology revolution) he said, "It is very easy to answer many of these fundamental biological questions; you just look at the thing!" [3]. At that time, everyone expected this would happen when the electron microscope was made 100 times better. We never got that improvement because the high-energy quanta used for observational resolution destroys the biology unless it can be preserved through techniques like those used by X-ray crystallographers. Henderson summarized in 1995 that this would be true regardless of whether neurons, electrons, or X rays are used as the illumination [4]. A different approach, however, had been developed by the 1970s with advances in computerized processing to detect low-energy quanta signals acquired over extended time periods. The new MRI was able to accomplish both body scans and molecular analysis, with functional MRI coming online in 1995. AFM, meanwhile, had been developed in the 1980s, with commercial devices proliferating in the 1990s. In 1994, Rugar reported force detection of nuclear magnetic resonance using MRFM. While medical MRI looks at groups of at least 1 trillion proton spins, the IBM researchers have detected the much fainter signal of a single electron spin. They also demonstrated imaging with 25-nm resolution, about 40 times better than the best conventional MRI-based microscopes.

The Advantages and Features of the Microscope

The MRFM uses strong field gradients to create the environment in which NMR pulse techniques can be interpreted. The central feature of an MRFM is the mechanical microscopic cantilever. An electric current induces a field within the molecule with which resonant interactions between a magnetic tip and the sample can be measured using NMR techniques. This is like holding two bar magnets and feeling the polar forces without actually touching them together. At a few attonewtons (10^{-18}), the force exerted by an electron spin on an MRFM cantilever is a million times weaker than the forces encountered in atomic force microscopy. By tuning an oscillating high-frequency magnetic field to the natural precession frequency of the spin being imaged, the spin's magnetic

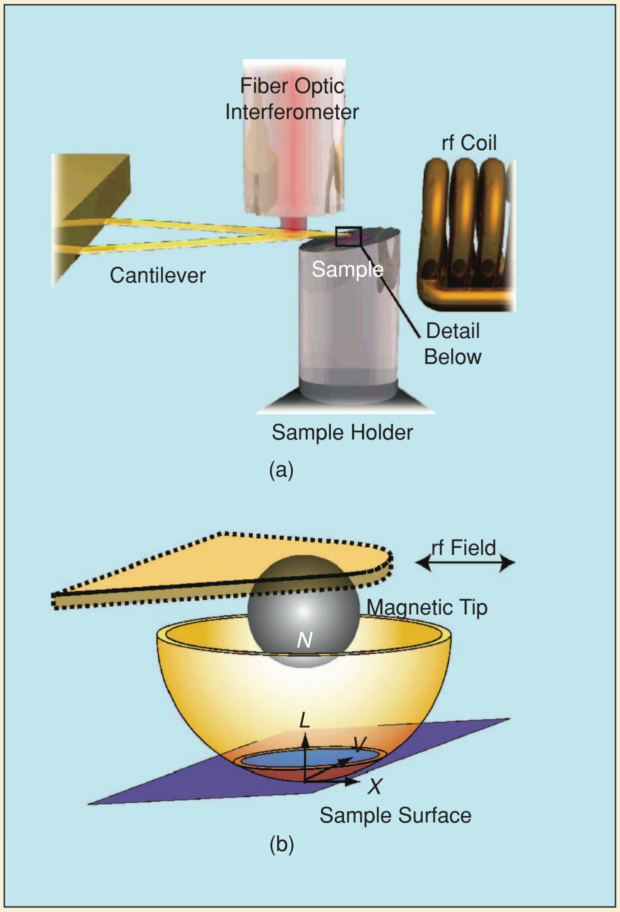


Fig. 1. An MRFM apparatus resembles a conventional magnetic force microscope with the addition of a rf induction coil. The magnetic tip induces a high-gradient polarizing field in the sample, and the rf coil imparts an amplitude-modulated flux density that periodically flips resonant spins. This modulates the net spin-gradient force, exciting a vibration of the cantilever whose displacement is detected by a fiber-optic interferometer. A conventional piezoceramic scanning tube moves the sample to create a 3-D force map providing the raw data for image reconstruction.

orientation flips back and forth as the cantilever vibrates. The flipping of the spin causes a detectable change in the cantilever's vibration frequency. The IBM researchers detected the signal of a single electron spin. (See Figure 1.)

MRFM results generally improve with sharper tips and smaller cantilevers. This directly relates to the sensitivity measure in the signal-to-noise ratio (SNR), but design tradeoffs emerge at the standard quantum limit. A stiffer cantilever, for example, offers better control over noise but requires a stronger signal to move it. Likewise, softer, lighter cantilevers are capable of sensitive detection but are buffeted by the process and measurement noise. Current experiments are conducted in a vacuum of around 10^{-5} torr, and IBM invested nearly US$1 million to reduce thermal noise with an Oxford Instruments millikelvin dilution refrigerator. Besides the cost and the fabrication facilities, you need the talent of someone like Dan Rugar who Sidles calls, "the Enrico Fermi of MRFM."

Cantilever fabrication continues to improve from a broad range of interests beyond MRFM, and this is one reason why the University of Washington (UW) group is focusing on digital signal processing (DSP) for systems integration. Because the cantilever is construed as a mechanical oscillator, sensitive detection depends on lower resonance frequencies that match with the spin precession, and these "Larmor frequency" cantilevers are what the UW group is testing before they are even yet available.

Image reconstruction from atomically mapped MRFM signals, however, poses a significant challenge for MRFM systems. Because the resonance occurs in a strongly curved shell that extends beyond the scan range, the geometry inherent to MRFM requires novel reconstruction methods similar to routine NMR. The scale and the curvature of the gradient complicate Fourier techniques that work well in conventional MRI. Despite these difficulties, the application of iterative reconstruction in the IBM experiments demonstrated roughly 25-nm resolution.

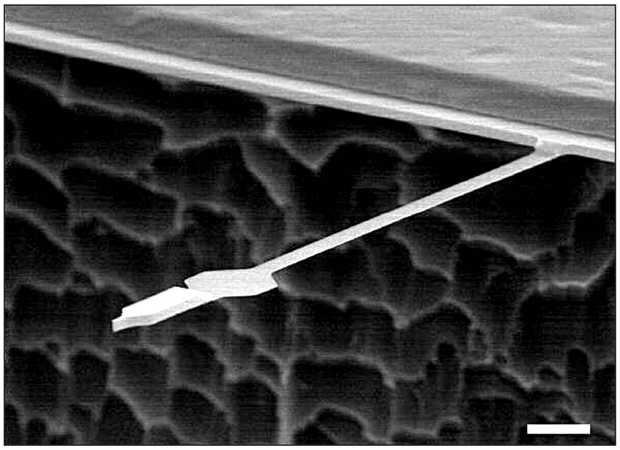


Fig. 2. Mass-loaded cantilever for single-spin MRFM [5]. Scale bar is 10 μm. (Image courtesy of IBM.)

Measurement noise corrupts state information, whereas process noise serves to drive cantilever motion.

Improvements to reduce the experiment time from days to hours are also needed, and these will develop as advanced cantilevers become available. Cantilever fabrication is undergoing an explosion of interest at nanotechnology facilities, where exotic designs promise devices with the lowest-possible quality (Q). (See Figures 2 and 3.)

The Q factor applies to an electrical circuit component that is inversely related to the fraction of the energy in an oscillating system lost in one oscillation cycle. Q is inversely related to the range of frequency over which the system will exhibit resonance. It affects the SNR, because the detected signal increases proportionally to Q while the noise is proportional to the square root of Q. The Q of a coil will depend on whether it is unloaded (no sample) or loaded (sample).

Designs in Controller Systems and Cantilevers

To be ready when new hardware is available, the University of Washington is applying an emulation design technique often used in manufacturing. Their approach is to emulate and test design parameters at the standard quantum limit (SQL) prior to full systems integration. They focus on controller systems while other groups pursue fabrication techniques for exotic cantilevers and cooling systems that operate below 10,000. Teaming with researchers at Cornell University who will build the cantilevers and at the University of Michigan for image reconstruction algorithms to manage the volumes of data they hope to deliver, the UW team has a three-stage process designed to deliver a working device. This maximizes their advantage in control expertise to prove that MRFM can be used to map three-dimensional (3-D) coordinates atomically when specified hardware and reconstruction algorithms are ready. Tom Kriewall recently published the results of a cantilever controller with a closed-loop, real-time digital signal processor [7].

The University of Washington is trying to test the principled limits at the standard quantum limit in channel capacity, rather than SNR, as the metric for new designs. This will establish the value for a wide range of designs, each of which has the potential to map samples with atomic resolution. The team includes signal-processing wizard Al Hero, at the University of Michigan, who contributed algorithmic techniques for the IBM experiments, and John Marohn, in the Chemistry Department at Cornell, who is the first postdoc in MRFM to achieve tenure with an MRFM laboratory. Education of future quantum systems engineers will play a fundamental role in the UW effort to characterize MRFM for commercial applications and train the next generation to industrialize molecular imaging.

Optimal control of their DSP system, they discovered, is tightly linked to the deterministic latency of a complete system. These systems involve more than just the input sampling rate or the processor clock speed.

The cantilever basically functions as an oscillator with real-time cycles at 50 MHz. Acquiring signals at this rate is amazingly routine for current hardware, and latency wouldn't matter if the goal were just to qualify that oscillation. But control of the cantilever is like playing a guitar, Kriewall says, and you can't play if you have to wait too long for the signal to wind its way through the processor every time you pluck a string.

In fact, you can't begin to expect that any processor will return an output in the 20 ns range (the single-cycle period at 50 MHz). The down conversion process determines the routes that signals follow, and the latency for a broad spectrum of signals depends on getting signals into and out of the processor and moving them across the bus. In the MRFM device, the signal from the instrument is acquired via a 65 MHz 12-b A/D converter and fed to a quad VME PowerPC board for processing. The real-time measure of the digital signal processor from the input wraps-around in continuous calculation of the proper output signal by adjustments to the control voltage via a 12-b 200 MHz D/A converter.

A Pentek Model 4294 Quad G4 PowerPC VME board with two velocity interface mezzanine (VIM) modules for floating-point signal processing was chosen for this high throughput I/O application. In the cantilever control system, bidirectional first in, first out (FIFO) memories of the model 4294 determine the synchronization of the signal input to the mezzanine interface. The VIM modules decouple the processor and the streaming, 32-b parallel interface, with synchronous, bidirectional FIFO memories, or synch biFIFOs. Synch biFIFOs provide consistent timing and buffer input and output data from the processor with block transfers operating at up to 100 MHz. The parallel interface biFIFOs in the Model 4294 allow the mezzanine port to operate at its maximum of 100 MHz; this frees the engineers from having to program the processor to take or deliver data at just the right time.

The synchronous, serial interface supports two 100 Mb/s full duplex channels, and the control/status interface provides microprocessor-like access for reading and writing to memory mapped registers on the mezzanine for configuring and controlling module functions. Most applications use these configurable interrupt flags in the biFIFOs to perform automatic data transfers between the processor memory and peripheral devices, thus leaving the processor open to perform other, peripheral tasks. The MRFM application, however, needs the flags for deterministic clock synchronization, because the cantilever control very much depends on knowing where the cantilever was, as well as when it was that you looked at it.

The central feature of a magnetic resonance force microscope is the mechanical microscopic cantilever.

On the processor side, the biFIFO is loaded or unloaded whenever possible, usually at the end of a processing loop in block transfers of data that make the most sense for general application. Prudent real-time signal processing design techniques require that the processor task execution time for a block of data is (at least slightly) shorter than the time it takes to collect that block. In this way, the processor finishes all of its "homework" and waits (perhaps briefly) for the next data block to become ready. BiFIFO buffering embodies the ideal implementation of this approach.

In the feedback loop for cantilever control, these block transfers unfortunately include a latency limit on the signal acquisition of most interest. Joe Garbini, professor of mechanical engineering at the University of Washington, explains it with the analogy of controlling an automobile. Imagine driving a car with the window blacked out. Two cameras collect your data on the front of the car (the cantilever), which is transferred to a computer in the back seat (via the VIM modules to the processor). At regular intervals, the processor dumps that data to a monitor in the front seat (via the bus on the mezzanine) where you see what's in front of you. The signal of most interest is not contained in the compilation of signals acquired in the block transfer, however fast that might have been, but only in the last signal that tells you the most current position of the cantilever (or the car).

Such time-sensitive applications are driving the real-time operating system (RTOS) market for lower latency where 50 μs is currently the benchmark limit for most commercial systems. Improvements in custom applications have focused on the mezzanine routing where field programmable gate arrays (FPGAs) offer flexible configurations

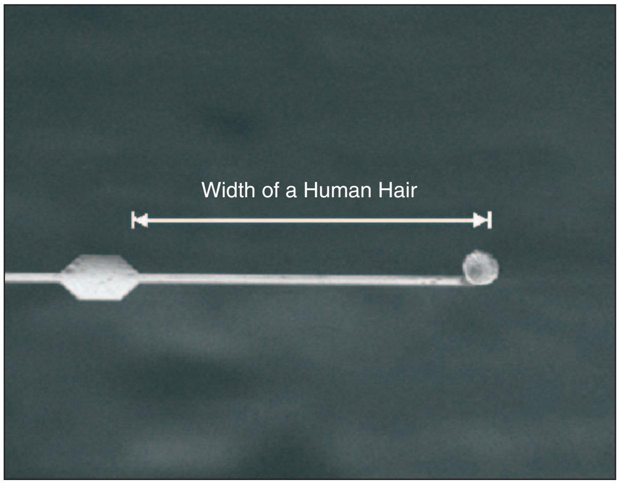


Fig. 3. SEM showing a 5-μm wide, 0.34-μm thick, single crystal silicon microcantilever. The sphere is a ~9-μm diameter nickel sphere, glued by hand to the end of the cantilever. (Image courtesy of Cornell University, John A. Marohn Group.)

for singularly deterministic tasks within larger systems. In the MRFM implementation, DSPCon was able to deliver a turnkey system using Vx Works software with 50–100 μs of deterministic latency measured within one board-loop from input to output. The goal for MRFM is to eventually operate I/O signals at around 5–10 μs.

In the current MRFM design, one mezzanine site hosts the model 6216 dual channel 65 MHz 12-b A/D converter with programmable-gain amplifiers, antialiasing filters, and digital downconverters. The analog cantilever sense signal is digitized by one of the A/D converters and then delivered to the processor board. The second mezzanine site is occupied by the model 6229 digital upconverter that includes two 12-b 200 MHz D/A converters with a dc to 80 MHz range. It accepts the synchronization signal input from the 6216 and one of the D/A converters generates the analog output control voltage to complete the loop. A Sun Blade 150 workstation serves as a software development workstation and is used to communicate with the Model 4294 via 100 Mb/s Ethernet (Figure 4).

This design allowed the investigators to test a novel method of heterodyne control for 25, 50, and 100 MHz signals. Heterodyne control is a method for reducing computational load and noise outside the passband and generating lock-in signals for online diagnostics, system identification, and adaptive control.

It is standard practice in control engineering to develop emulators and controllers in parallel, using each to debug the other. For the MRFM program, the main practical advantage is that when the advanced cantilevers are fabricated, the controllers to use them in MRFM experiments will be ready; this tactic improves the overall pace of research and development and allows MRFM cantilevers to be fabricated with good confidence that they will perform as specified.

Tom Kriewall, now a line-engineer at Intel, developed this technique for his Ph.D. dissertation.

> Heterodyning changes the center frequency of a narrowband signal without changing its bandwidth. Downconversion uses the heterodyne principle to generate two quadrature signal components at an intermediate frequency (IF) from a single high-frequency signal. These low-frequency components are referred to as the in-phase (X) and quadrature (Y) signals and are the same as those produced by a lock-in amplifier.
>
> Down conversion can be thought of as a translation between two domains, namely from the unheterodyned or lab frequency (LF) domain to the IF domain. Up conversion entails the opposite process by which a high-frequency LF signal is reconstructed from the IF signals. Synchronization between the down conversion and up conversion processes is essential for heterodyne control. [7]

In terms of latency specification for the DSP, this means that when the IF is retuned, it is acceptable for the filter to shut down for a period of up to 100 μs, provided that when the filter restarts, the phase response is specified. Future generations of the MRFM device will employ cantilevers that resonate at 1–10 MHz, and improvements will be necessary to match the carrier frequency with this natural frequency. Conventional control is inefficient for this purpose due to the Nyquist theorem, which requires sampling and data processing rates that are higher than the cantilever resonance frequency.

The IF domain cantilever model is useful for three purposes:

- easy development of the heterodyne controller
- computationally efficient simulation of high-frequency narrowband micromechanical oscillators, either open loop or under closed-loop control
- emulation of high-frequency narrowband micromechanical oscillators. This optimally combined heterodyne controller and estimator includes techniques applicable to any narrowband system.

The IF cantilever model is formed with the "hetero," or "many," signals of $x(t)$ and $y(t)$ for in-phase and quadrature position, combined with in-phase and quadrature force signals $g(t)$ and $h(t)$. Up conversion to in-phase force signal and the quadrature position signal in thermal noise and cantilever mass produce a force input $u(t)$ to the

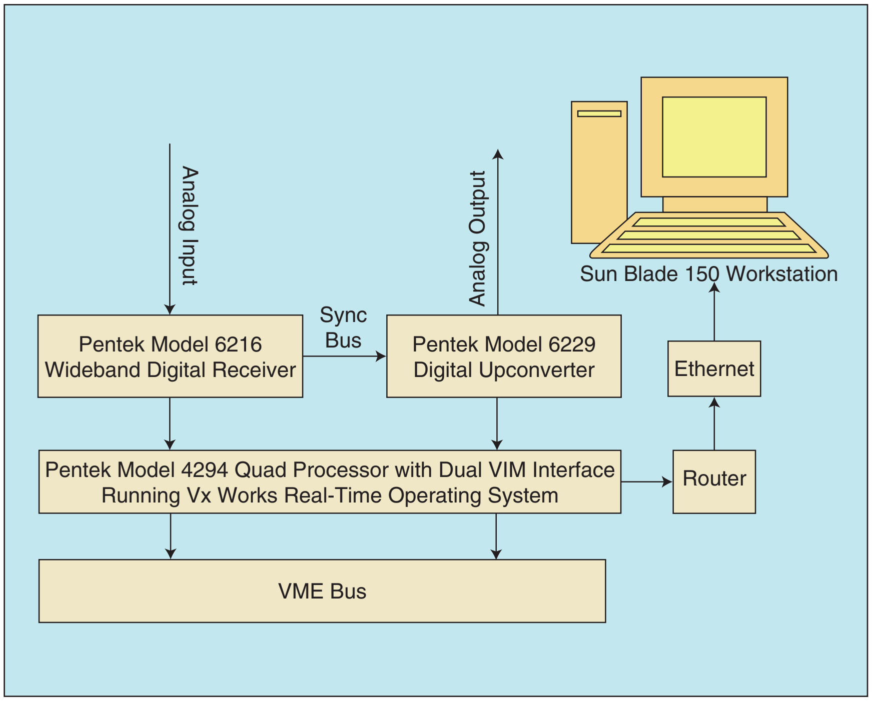


Fig. 4. DSPCon Pentek DSP system.

cantilever matched in down conversion of the position output within thermal noise as in-phase and quadrature position (Figure 5).

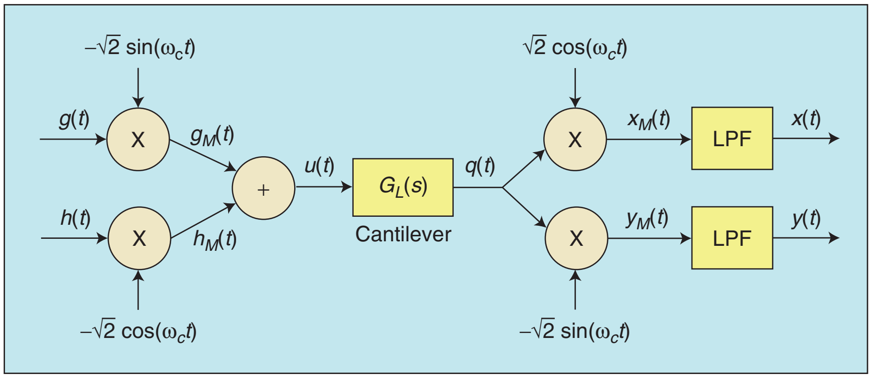


Fig. 5. A cantilever model.

This estimation model allows the control problem to be entirely cast in terms of IF domain quantities. The down conversion shifts the center frequency of the force signal $u(t)$ to dc, allowing the low-pass filters to function about dc the same as the resonance peak functions about the natural and carrier frequencies. Experiments by Kriewall show that when the IF domain signal is up converted to the natural frequency $w(n)$, the output from the LF and IF cantilever models are identical [7].

With equivalent frequency estimation, the separation principle allows IF domain control to be joined with the estimation between a downconverter and an upconverter forming the heterodyne control. Figures 6 and 7 illustrate the features added to conventional control with the heterodyne technique.

In general, measurement noise corrupts state information, whereas process noise serves to drive cantilever motion. A higher process noise results in a higher output signal, thereby reducing the measurement noise so that the measured signal more closely matches the actual cantilever position. This allows higher gains to be applied to the estimator with faster convergence of the estimated state to the true state. Driving the cantilever to a higher spring constant overrides thermal noise because the cantilever gets stiffer, and this increases the effect of the measurement noise.

As more aggressive control is applied, frequencies over a broader range of deviation from the origin are allowed to pass the low-pass filters. Increasing measurement noise similarly increases process noise as the IF estimator is able to track the higher frequency components.

MRFM illustrates how signal estimation is configured in a novel method of heterodyne control. Emulation and control is implemented in parallel crates so that each can be used to test and debug the other. This approach reduces the risk and cost of development because the specifications can be rigorously determined prior to ordering custom or off-the-shelf integrated systems.

Hardware latency places an upper limit on the allowable bandwidth of this closed-loop system. While the heterodyne controller acts on the slowly varying envelope signal where cycle delays can be tolerated, the total hardware latency must nevertheless be calculated or measured before a controller design can be applied. For a DSP system similar to the UW MRFM group, the design rule may be expressed as the mixer latency of the up and down conversion, plus latency in the biFIFO buffers due to asynchronous communication, less the system time constant where the system time constant is the time-rate of change for the envelope. This enforces an upper limit on the closed-loop bandwidth, but the system is robust to small deviations. Hardware latency also introduces a deterministic phase shift that can be managed in different correction schemes, and mismatch between the carrier and natural frequency must be monitored. Specifying the total hardware latency is an important calculation that must be specified for either custom or commercial off-the-shelf (COTS) purchases.

In this application, control of an emulated cantilever was proven at resonant frequencies of 25, 50, and 100 MHz. The design team can now confidently specify the systems integration knowing that the complicated controllers and the expensive cantilevers will work as expected.

Quantum System Engineering Development Incentives

The Department of Defense has announced the 2005 MURI Program award winners for their topic area #6, Advancement of Magnetic Resonance Force Microscopy to Single Nuclear Spin Detection. John Sidles will lead the UW team in one award, and P. Chris Hammel from Ohio State University will lead a collaborative team on a parallel effort. The Hammel team includes Michael Roukes at Caltech, who recently published heterodyne downmixing techniques for detecting the presence of zeptogram molecules with ultrasensitive

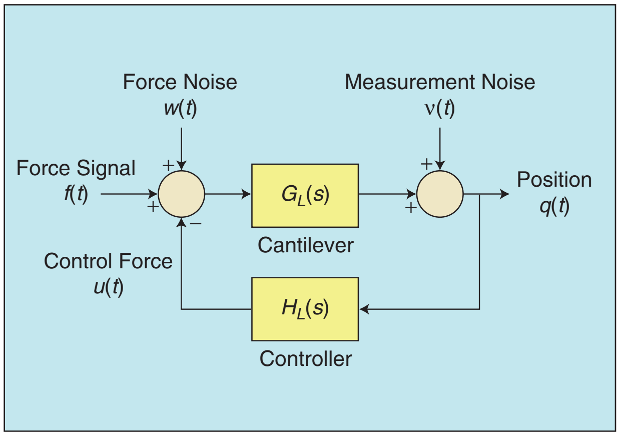


Fig. 6. A closed-loop MRFM cantilever system. Feedback control $u(t)$ is applied to the MRFM cantilever to broaden the spectrum of the position signal, thereby facilitating practical MRFM imaging.

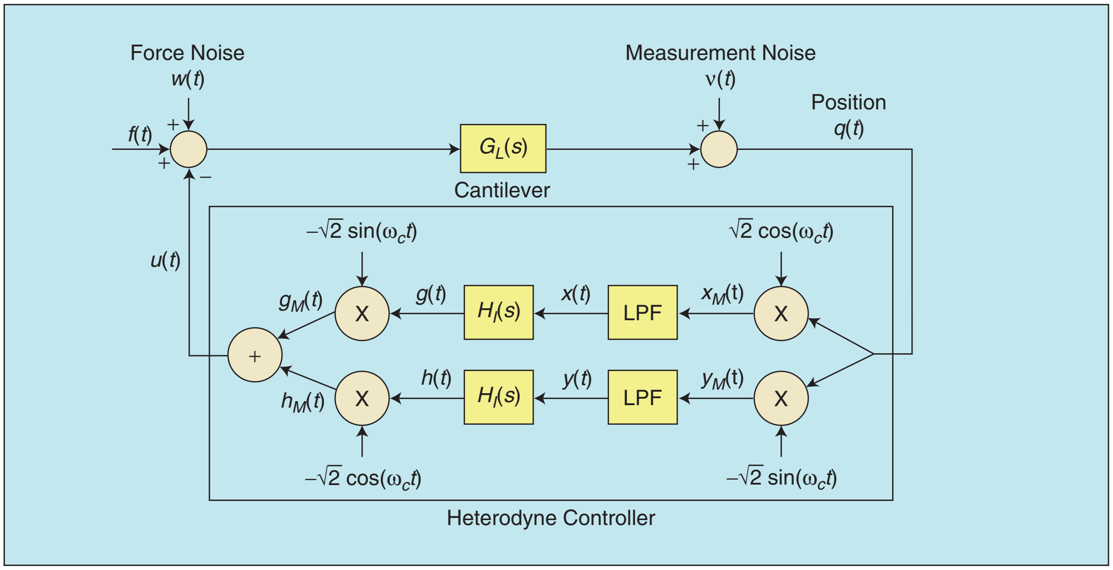


Fig. 7. Closed-loop cantilever system with heterodyne controller.

cantilevers. Both Roukes and Hammel have made significant contributions to the national Spintronics and Quantum Computing programs. In addition, the Hammel team has joined with Raffi Budakian at the University of Illinois, Urbana-Champaign, coauthor on the Rugar single-spin detection article. The UW and Ohio teams are sharing results on their MRFM projects, IBM is a consult to both groups, and collaborations continue with teams at other universities and national laboratories, as well as international groups interested in MRFM, the Army Research Laboratory, the West Point Academy, and emerging small and large businesses. MRFM has generously received funding from the NIH, NSF, and the Department of Commerce. Roadmaps are drawn among the interested parties to advance MRFM to a program level of commercial and national importance equivalent to other engineering programs testing how the quantum limits might be understood and specified in new devices.

Spintronics, for example, promises to advance computer storage and retrieval using sensitive manipulation of the magnetic spin properties. Superconducting quantum interference devices (SQuIDs) are being used to control quantum bits in computing by superposition, and the Laser Interferometry Gravity Observatory (LIGO) is attempting to map the stars by detecting gravity waves in quantum cavities. Despite the broad scale of these technologies, they all acquire signal averages over time with optimized low-quanta energy. Control theory dominates the manipulation of these systems, and engineering standards are needed as a foundation for commercial designs at the quantum limit. The UW team believes these standards will be measured in channel capacity limited by process and measurement noise. MRFM has the comfortable advantage of working in the range of about 40 dB above the SQL where these quantum systems can currently be advanced in graduate studies.

References

[1] D. Rugar, R. Budakian, H.J. Mamin, and B.W. Chui, "Single spin detection by magnetic resonance force microscopy," *Nature*, vol. 430, no. 3, pp. 329–332, 2004.

[2] J.A. Sidles, "Noninductive detection of single-proton magnetic resonance," *Appl. Phys. Lett.*, vol. 58, no. 24, pp. 2854–2856, 1991.

[3] R.P. Feynman, "There's plenty of room at the bottom," *J. Microelectromech. Syst.*, vol. 1, no. 1, pp. 60–66, 1992.

[4] R. Henderson, "The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules," *Quart. Revs. Biophysics*, vol. 28, no. 2, pp. 171–193, 1995.

[5] B.W. Chui, Y. Hishinuma, R. Budakian, H.J. Mamin, T.W. Kenny, and D. Rugar, "Mass-loaded cantilevers with suppressed higher-order modes for magnetic resonance force microscopy." in *Proc. Tech. Dig. 12th Int. Conf. Solid-State Sensors Actuators (Transducers'03)*, Boston, MA, June 8–12, 2003, pp.1120–3.

[6] Department of Defense, Office of Naval Research, Science and Technology Division, Industrial and Corporate Programs, Multidisciplinary Research Program of the URI (MURI) [Online]. Available: www.onr.navy.mil/02/baa/docs/04_021.pdf

[7] Kriewall II, T.E., "Heterodyne digital control and frequency estimation in magnetic resonance force microscopy," Ph.D. dissertation, University of Washington, 2004.

Doug Mounce (mounce@u.washington.edu) is the manager for the MRFM program at the University of Washington. He also manages the Department of Orthopaedics' Regenerative Orthopaedics Program and worked for a number of years in the University of Washington's Office of Sponsored Programs. He has the operational role of turning vision into action and thinks he has the greatest job in the world. His interest is historically centered on the people of MRFM, their knowledge, their instruments, their collaborations, and their goals.