Translational Engineering: Best Practices in Developing MEMS for Volume Manufacturing

Alissa M. Fitzgerald,* Keith M. Jackson, Charles C. Chung, and Carolyn D. White

A.M. Fitzgerald & Associates, LLC, 700 Airport Blvd. Suite 210, Burlingame, CA 94010, USA

New MEMS devices are usually invented by academics having the goal of advancing the technology and knowledge in their field. Along the way, market opportunities are often found, and then a new goal arises, to commercialize the technology. However, the initial prototype was never engineered to meet this new goal. Before a new MEMS device can be commercialized, it must be reengineered and adapted for the volume manufacturing environment. In this paper, we describe our method of "translational engineering", developed over the past 15 years, to translate proof-of-concept prototypes made by academic inventors into robust, advanced prototypes that can be successfully transferred to production.

1. Developing MEMS for Volume Manufacturing: Overview

1.1 From research project to commercial product

Many new MEMS devices, whether sensors, actuators, or passive microstructures, are invented and initially developed in a university or a government-sponsored laboratory. In that setting, researchers focus on demonstrating new physics of operation, or enhancing performance capabilities using new materials and methods, and build their first "proof of concept" prototype. The researchers create these prototypes using the tools available within their own laboratory, typically much older models that had been donated or purchased used. Often, because of limited equipment or budget, manual fabrication steps may be used. In a research project, a successful prototype is defined as one that provides sufficient insights and data to publish a peer-reviewed journal article, so these fabrication limitations are acceptable.

MEMS (or semiconductor device) research often inspires entrepreneurial ambitions, and many new companies have been formed on the basis of a founder's Ph.D. dissertation. However, one cannot simply send a successful research prototype straight to a foundry for commercial manufacturing. The research prototype had been developed solely for academic purposes, so it has several deficiencies for commercial manufacturing: (i) the relationship between process tolerances and device performance is not yet fully understood, (ii) the process may involve the

---

*Corresponding author: e-mail: amf@amfitzgerald.com

use of machines, materials or methods not commonly available in production facilities, and (iii) the design and process have not yet been optimized for items crucial to a commercial product, that is, packaging, testing, high yield, and low cost.

Further advance of a technology created in a research environment requires specially focused development work. Sometimes referred to as "design for manufacture", we call this work "translational engineering", because the original intent of the inventors must be interpreted and translated into a design that can be manufactured in volume (thousands to millions or billions of units per year). This work is needed for MEMS especially since the design and fabrication process of a MEMS device are so interdependent that small design changes will impact the process flow and vice versa. Depending on the complexity of the device design and its process, translational engineering can span years and consume millions of dollars, before commercial production can begin. It is a necessary and unavoidable step in MEMS development. Many MEMS startups have failed because the time and funds needed for translational engineering were badly underestimated.

In this article, we will describe our methodology and best practices, developed over the past 15 years and more than 160 client projects, for translational engineering of new MEMS designs in order to prepare them for successful volume manufacturing.

1.2   Understanding the manufacturing environment and economics

The goal of translational engineering is to deliver a MEMS design and process flow to a production fab for manufacturing. To sharply focus such development efforts, one must first understand and appreciate the volume manufacturing environment.

Wafer fabrication facilities ("fabs") are complex factories whose construction costs a minimum of $100 million for MEMS production and a minimum of $2 billion for state-of-the-art semiconductor production. These costs, and significant recurring operating costs, can only be justified by operations focused on high manufacturing throughput, 24 h/7 d a week operation, and equipment utilization rates as close to 100% as possible.

Foundries (contract manufacturing fabs) therefore seek customers who will buy large quantities of wafers per year. Minimum order quantities of 5000 wafers per year are common for high volume MEMS foundries producing 200-mm-diameter wafers. Even smaller foundries, producing 150-mm-diameter wafers, may require minimum order quantities of 500 wafers per year, with 100 wafers per year being the absolute minimum.

Fabs typically run dozens of different products through their facility. In some cases, fabs run both CMOS (semiconductor) and MEMS products, each of which have distinct process flows, through the same facility. Managing so many groups of wafers moving along different paths through the fab and keeping the tool utilization high require complex and detailed tool scheduling. Each tool will have a queue of wafer batches waiting their turn, and disruption of that queue or the tool (for example, to conduct experiments) will cause cascading schedule problems. Owing to this complex operating environment, fabs strongly favor producing MEMS that will be compatible with their existing tools and processes.

The foundry business model demands the selection of customers having the lowest risk processes at the highest possible profit margin in order to derive the most profit possible from

the fixed production capacity of a facility. While a foundry might consider, for strategic reasons, accepting a customer with a new type of design or process, they will very likely charge higher prices to compensate for the anticipated disruptions to their existing operations. Any development work undertaken by the foundry requires special attention from foundry engineers, which will be charged to the customer (as nonrecurring engineering fees). The foundry will also want to retain rights to any new process intellectual property (IP) developed. If a customer's process is deemed too early stage or too different from core processes, the foundry will likely decline the business outright.

With that perspective, one can better appreciate that a proof-of-concept prototype is too fragile to go straight to a production facility. The translational engineering work to be carried out must be focused on ruggedizing the technology for the demanding production environment; the MEMS design and process flow must be engineered to require a minimum of human intervention during fabrication, have process tolerances that are comfortably met by existing fab equipment, and for each process step, have well-defined pass/fail criteria, which can be easily inspected using common metrology equipment.

In summary, a totally new prototype must be designed and built. This advanced prototype is what will eventually be transferred to a foundry for production.

1.3   Preparing for manufacturing

The advanced prototype demonstrates readiness for manufacturing. In addition to the functionality demonstrated by the earlier proof-of-concept prototype, an advanced prototype must also have the following new attributes.
- A model of how process tolerances affect device performance
- A process flow and mask layout that can be executed in a production fab
- A design that considers downstream packaging, testing, and system integration needs
- A fabrication cost that allows adequate profit when sold in a given market

The translational engineering work needed to explore and develop these advanced prototype attributes is described in Sect. 2 below.

2.   Designing Advanced Prototypes

2.1   Developing parameter sensitivity models

A device technology is not fully mature nor manufacturable until one understands how all process parameters contribute to its proper function. In other words, how sensitive is the device performance to variation in each process step? Knowing the parameter sensitivities enables both implementation of inspection on that process step and establishment of pass/fail criteria to screen out wafers whose process variations will cause device failure.

For example, film thickness is one of several parameters that affect the stiffness of a membrane device and its resonant frequency. How thick or thin could that film be before the variation in stiffness impairs overall device performance? Membrane stiffness is proportional to the cube of film thickness. If the required device performance depends on controlling

membrane stiffness to within ±10%, then the film thickness must be controlled to within +3.2 and −3.5% (the cube roots of 1.1 and 0.9, respectively). If a deposition tool cannot repeatedly perform within those thickness tolerances, then the process will depend on luck (random variable) to achieve the correct film thickness, and will therefore have poor yield.

Exploring and understanding parameter sensitivities is best done using simulation. The simulation environment allows one to explore the interaction of many design parameters much faster and more cost effectively than by building and measuring actual devices.

First, an adequate model of the device physics must be created. The model does not require precise material properties data nor does it need to look exactly like the finished device; it must, however, capture the fundamental physical behaviors of the device. At this stage of the development, we seek to understand how relative changes in input variables affect device performance, not to calculate absolute values with precision.

Often, a lumped parameter model (such as the equations for a mass-spring-damper system) is sufficient to elucidate the sensitivity to major process variables. Such a model could be implemented on an Excel spreadsheet or a Matlab script and used to quickly identify the most sensitive parameters and their approximate range of acceptable tolerances. Once first-order behaviors and sensitivities are well understood, then a more advanced model could be created by finite element analysis (FEA) simulation to study the more subtle parameter interactions. For example, FEA is well suited to explore interactions with 3D geometries. FEA models can be time-consuming to build and verify, so engineering judgment must always be applied to determine the appropriate level of detail in a model. The ideal FEA model contains only enough features to correctly simulate the critical physical behavior and no more.

Data and insights gained from parameter sensitivity modeling must inform process integration and design layout. Typically, several iterations are needed between modeling and process integration before convergence to an advanced prototype design.

2.2 Process integration and mask layout for manufacturing

Designing an advanced prototype requires creating a process and mask layout that can eventually be executed by a production fab. The following factors are important when translating a proof-of-concept design.
• Selecting processes compatible with those at production fabs
• Engineering the device design to function within reasonable process tolerances
• Having clear prototype performance goals in order to guide process and design tradeoffs

For smooth commercialization, it is essential to create an advanced prototype using processes commonly found in production fabs. Any chemicals, photoresists, or tools needed for the process must already be commercially available. Processes should not require individual wafer-by-wafer tuning, nor any manual steps. All materials and chemicals must be compatible with the types of foundries to which the product could eventually be transferred. For example, if the likely manufacturer will be a CMOS foundry, then materials such as gold or processes such as KOH etching, both of which contaminate CMOS devices, cannot be used.

Even with processing of very large volumes of wafers under stable conditions, all

manufacturing processes have some random variation that will cause a plus or minus tolerance on dimensions and material properties. An advanced prototype design must be engineered to work within the limitations of available processes. This requires a deep understanding of how typical manufacturing processes perform, and then creating a design that can accomodate those process imperfections. Creating designs that can succeed within typical process tolerances will maximize the selection of candidate foundries, which in turn will help get competitive pricing for volume manufacturing.

When developing an advanced prototype, the goal should not be perfect performance but making sure the device will function. There might be one step where tight process tolerances may be required, but it is always worth considering if sacrificing a certain performance will allow a wider tolerance and therefore a higher overall chance of creating a working device. Test data from an imperfect device is very valuable, because it will provide useful data for tuning the models and the design, and for identifying further process optimization. A second, subsequent prototype could always be used to further improve the design and process. The opportunity to learn is greatly diminished if a prototype fails to demonstrate even basic functionality.

Interactions and tolerances between the registration of different mask layers must also be carefully considered. The results of this analysis will eventually help to establish design rules for future device design. Minimum linewidth or spacing between the features on each layer is defined by the lithography variation and etch accuracy. The minimum overlap or spacing required between layers is defined by a combination of lithography variation and layer-to-layer alignment accuracy of the exposure tool.

Typically, misalignment errors and lithography variations are considered to be normally distributed random errors. This enables one to calculate an overall expected error from accumulated tolerances by adding the sum of the squares of each contributing error and then taking the square root. An advanced prototype's layout should be designed accounting for a realistic lithography error "budget".

In MEMS, process and design are inseparable. While considering tradeoffs between the two, the big picture in business and technical goals must always guide engineering choices. Whether the technology is being commercialized by a startup company or a Fortune 500 company, prototypes must always demonstrate capability in order to be further funded. As different processes or designs or layouts are considered, they should be evaluated and guided by the goals of what the prototype must eventually demonstrate. Choices should always be conservatively made to ensure that it will be possible to yield some working prototypes, even if they have less-than-ideal performance. An overly ambitious, high-risk prototype that is designed idealistically for a perfect outcome but ultimately fails to work in practice is much less useful.

2.3   Designing for testing and data gathering

Testing is essential for evaluating a prototype and designing it for manufacturability. In a wafer manufacturing process, there are three opportunities for testing: in-line testing, back-end-of-line (BEOL) wafer testing, and package testing. In-line testing is done during the processing

of the wafer. BEOL testing is done after the wafer process is completed. Package testing is done after the devices have been singulated from the wafer and mounted into packages.

There is an important tradeoff between timeliness of information versus quality of information. Information obtained during the manufacturing process can identify defective devices or wafers early on, when less money has been spent. The highest quality information comes at the end of manufacturing, during the final package test when the device is tested under conditions of realistic use; however, this is also the point at which the device is most valuable.

A test plan is essential, and for an advanced prototype, these different testing points and their tradeoffs and relative costs must be considered. Planning is required to ensure that an advanced prototype will incorporate any special features or structures needed to facilitate testing, and that sufficient test wafers will be available for any destructive tests.

In-line testing may be nondestructive or destructive. An example of nondestructive testing is midprocess electrical probing to verify that a process step had been correctly completed. A typical probe test would measure the resistivity between two contact points after a metal deposition step. After testing, the wafer would resume its process flow.

A common example of destructive in-line testing is scanning electron microscopy (SEM) of a device cross section. A wafer would be removed from the process batch, and to expose its cross section, it would be cleaved or sawed. To facilitate this type of inspection, test structures must be designed so that their cross sections can be easily cut, and multiple structures would be arranged on the wafer so that they can be exposed by a single cut. Ideally, this type of test structure would also be closely spaced so that a single SEM observation could image multiple structures simultaneously.

After a wafer is completed, the test data quantity and quality increase owing to the use of automated wafer electrical probing. In automated probing, every device on a wafer may be electrically stimulated and measured to accumulate a large volume of data. Often, high-level device functions can be evaluated by wafer probing. For example, an accelerometer's sensitivity may be estimated by measuring the slope of its capacitance–voltage (C–V) response. An input voltage is applied to an accelerometer, and then its output capacitance measured. The capacitance changes may be on the order of femtofarads or smaller, and they must be measured in the presence of parasitic capacitances that are typically orders of magnitude larger. Designing an advanced prototype to facilitate these tests could include methods to electrically isolate the structures under test, such as surrounding the devices with a Faraday cage of known voltages, or arranging differential readout to isolate the test signal.

Once a chip is packaged, many more device behaviors may be investigated and characterized. Some behaviors can only be evaluated in the package, for example, the temperature response of a device. The thermal response of a MEMS device is strongly dependent on the packaging, since the package is usually made of a material that has a coefficient of thermal expansion different from that of silicon. The effects from thermally induced strains can only be measured and characterized after the device has been packaged.

An advanced prototype must, at minimum, have a bond pad layout and footprint compatible with the intended package. This means that before wafer layout and processing occur, some careful thought must be given to which package will be used, and which types of tests will eventually be performed on the package.

2.4 Designing for package and system integration

The primary function of a package is to mechanically and environmentally protect the MEMS die, while providing electrical contact and enabling any input or output stimulus to reach or exit the die. Package design is rarely considered thoroughly during academic research, because at that stage, bare die testing is sufficient for gathering data. Package design and assembly, however, are always critical parts of translational engineering development and are required for successful commercialization, because the MEMS must function when installed in a circuit board or other system.

The package is an expensive part of MEMS devices and can account for more than 70% of the total manufacturing cost. One reason for this high proportional expense is that while in wafer form, the manufacturing costs are spread among all the dies on the wafer, which may number in the tens of thousands. Once the MEMS dies are singulated from the wafer, however, further manufacturing costs instead scale on a per die basis. Yield loss at the package stage is therefore quite expensive.

Another reason for packaging expense is that MEMS devices are partially mechanical in nature, and the mechanical environment that the package presents to the die must also be well engineered. MEMS devices respond to mechanical changes, such as the stress in the membrane of a pressure sensor, or a dimensional change, such as a capacitive gap in an accelerometer. These changes may also be induced by mechanical forces originating from the packaging. As a result, the package can act as a strong error source for the device.

There are several approaches to managing the package's effect on device performance. One is to prevent or minimize any mechanical input from the package. This requires the selection of stiff packages that do not flex or bend easily, which usually translates to large and expensive packages. An alternative approach is to mechanically isolate the MEMS device from the package or even within the die. The MEMS die may be attached to the package in a way that does not transmit package stresses to the die, such as using a soft polymer to attach the die. Additionally, a clever device design could isolate the MEMS device from external influences. A common isolation technique for resonators, for example, is to collocate the resonator's anchors at a single point to minimize mechanical cross-talk.

The packaging also provides an opportunity to add value to the MEMS when developing an advanced prototype. For example, a design that accommodates specialized packaging would enable a MEMS sensor to operate inside the human body, such as the 1 French (0.33 mm diameter) guidewires that enable in vivo blood pressure measurements within the cardiovascular system.

The challenge of MEMS packaging increases with the complexity of the physical input or output with which it interacts. For example, packaging for a MEMS microphone must allow an acoustic wave to reach the MEMS sensor, while simultaneously protecting the microphone from factors such as water, particles, mechanical shock, and temperature variation. In this case, the MEMS die itself could be engineered to withstand some of these adverse environmental effects. For example, when developing an advanced prototype, one might also experiment with using a water-repellent coating to protect the surface of the microphone.

2.5   Cost considerations

While keeping in mind all of the attributes described above, the advanced prototype must also demonstrate that it can be manufactured at a reasonable cost, consistent with its business model. Although increased wafer volume (thanks to economy of scale at the foundry) may eventually decrease production cost significantly (to 25–50%), the advanced prototype must immediately demonstrate the correct order of magnitude cost.

For example, if the business model requires a cost of $3/unit, the advanced prototype should cost no more than $10/unit, with the expectation that volume production will eventually reduce the cost from $10/unit to $3/unit, over time. An advanced prototype that costs $100/unit will never be able to reach $3/unit; the economy of volume manufacturing cannot bridge that large a gap.

One of the major drivers of cost is chip size. Foundries charge per wafer, not per chip. Decreasing chip size therefore increases the number of chips per wafer and is therefore a big lever to reduce cost. The advanced prototype must be made as small as feasible.

Other significant drivers of costs are use of single-wafer process steps that cannot be easily reworked, for example, deep reactive ion etching (DRIE) and wafer bonding. Both processes have low throughput (wafers processed per hour) because only a single wafer is processed at a time in the tool, and the process itself can be slow (up to hours). Furthermore, process deviation or failure in either of those processes will result in loss of the entire wafer. Use of expensive start materials, such as silicon-on-insulator (SOI) wafers or high-resistivity float-zone silicon wafers, may also contribute significantly to the cost, and should be eliminated or substituted, if the device physics allows.

Finally, device yield (the percentage of good chips per wafer) is also a strong driver of chip cost. Engineering a design and process flow to fit comfortably within fab process tolerances is fundamental to keeping the yield high. Designs that depend on tight process tolerances will yield poorly, and will, by definition, be expensive.

3.   Fabricating Advanced Prototypes

3.1   When to use a development fab versus a production fab (or foundry)

Once the advanced prototype has been designed and is ready to be fabricated, a common question is "Where is the best place to fabricate it, at a development fab or a production fab/ foundry?"

The answer depends on several key criteria.
• How similar the advanced prototype is to existing production at a fab
• Whether the prototype needs strict process control to succeed
• Whether the customer has the expertise and time to properly manage the fab
• Funds and timeline available

Development fabs are smaller facilities (<20000 sq. ft.) that may have production-quality equipment, but whose main business model focuses on process research and development. A development fab is best for fabricating advanced prototypes when the device physics and its

interaction with process tolerances are still being explored or when flexibility is needed to develop process recipes and run multiple short-loop experiments. Being oriented towards R&D, development fabs generally have more engineers available to develop processes and work with customers. Overall, prototyping will cost less than in a production fab, owing to its smaller, and often older, depreciated facility; however, tradeoffs may exist in quality control.

Advanced prototypes may be successfully fabricated at a production fab or foundry if the device physics is well understood, very little process development is needed, and the foundry is already experienced with the device technology. At production fabs, processes are run by operators, not engineers, and tool priority is given to wafer production, not experiments, so prototypes needing more attention are better served at a development fab. Advanced prototypes that require a high degree of process control or uniformity will benefit from a production foundry that has the most up-to-date tools and quality control.

Customers working with foundries should have a knowledgeable MEMS engineer dedicated to working with the foundry to monitor progress and provide feedback on process data, because the foundry engineers will likely be too busy to provide detailed attention to a smaller project. Finally, customers should expect higher costs and longer timelines when running prototypes at a production fab or foundry, for the reasons described in Sect. 1.

Fabrication costs and timeline depend strongly on process complexity and whether custom wafers are being used. In general, fabricating advanced prototypes at a development fab will cost a minimum of $200000 and take four to six months, and at a production fab, will cost at least $500000 ($1 million if on 200 mm wafers) and take close to a year.

3.2   Short loops and planning

Risk mitigation during prototype processing involves gaining information before, during, and after fabrication. Identifying risks, and thus items to be measured and monitored, requires having context of the overall goals for the development effort and an understanding of which processes and specifications are critical to successful device function.

"Short loops" are test runs of a subset of process steps to investigate unknowns, such as the tool recipe settings to produce the required material parameters, or to test the interactions between two process steps. Short loops are generally implemented to reduce the risks of the most critical or unknown process details. Short loops should be run before design and mask layout work is completed, so that all data gathered can be immediately applied to design improvements and process integration.

When processing has begun, sometimes a "look-ahead" test wafer may be needed to validate a single process step prior to committing the batch of device wafers. The test wafers could be processed one or two steps ahead of the device wafers to allow quick verification that the process step will succeed. A look-ahead wafer is less expensive than a short-loop experiment because it would be done in parallel with the device wafers.

Early analysis of process integration is essential to identify the top risks and unknowns, and must inform fabrication planning. Part of a good fab plan is making sure there will be enough wafers to cover all the short loops, look-aheads, and destructive tests that will occur throughout the fabrication of advanced prototypes.

4.   Transfer to Foundry

4.1   Finding the right foundry

Once an advanced prototype design is stable and a company has a foreseeable demand for its product, it is time to transfer the technology to a production fab or foundry. Choosing a partner for volume manufacturing is one of the most critical decisions a company will make, because of the high cost ($ millions) and time (years) it takes to engage with a foundry and build up production there. A company should choose its foundry with the same level of care and time one applies to choosing a business partner or investor.

The foundry's capabilities must meet both the technical needs of the product as well as the company's business needs. Time must be taken to visit many candidate foundries and apply due diligence to verify that all process and business needs can be met. As in any business partnership, the executives and the engineers from both the foundry and the company must build trust in order to have a healthy, long-term working relationship. MEMS production is complicated and will always encounter unexpected hazards and setbacks. The partners must be prepared and determined to work constructively together.

4.2   Transferring an advanced prototype to a foundry

Once a foundry relationship is established, the technical team from the customer's side must transfer the technology to the foundry's process engineering team. Because tool sets and methodologies differ between facilities, it is not as simple as just sending over the advanced prototype's mask files and runsheets.

The customer's team needs to translate its work so that the foundry team truly understands the technology. Mask files, runsheets, and 3D process visualization software can provide much of the information, but not all. The foundry team must be taught why certain process steps or mask design features were chosen and their significance; what had been attempted earlier and did not work; which process tolerances (for every step) are essential to success and which will cause device failure; which features and test structures to measure and monitor in-process and how to correctly interpret that data; and what to measure and inspect postprocess. As of the time of writing, there is no software tool that can completely transfer this detailed knowledge of MEMS devices in an automated manner. Ultimately, the team that engineered the advanced prototype needs to work directly with the foundry team, for weeks to months, in order to fully complete the transfer.

Companies should start the foundry selection and transfer process at least one year prior to the desired start of manufacturing, and at least two years prior to the start of volume production (>1000 wafers per year). In the latter case, the foundry needs additional time to "qualify for production", which involves running many wafer batches to tune processes for high yield and automated fabrication. Companies should plan to spend at least $500K per year for a transfer to a 150 mm wafer fab and at least $1M per year for a transfer to a 200 mm wafer fab.

5.   Summary

A new device or technology initially created in a research environment needs "translational engineering" to prepare it for the rigors of the manufacturing environment. A different type of engineering focus is needed to interpret the inventors' intentions and to engineer an advanced prototype to be compatible with production fab tools and methods. In developing and fabricating the advanced prototype, a deeper understanding of the device technology is gained. It is in this deeper understanding where additional competitive advantages and intellectual property may be discovered, and where successful volume production is always achieved.

About the Authors

Alissa M. Fitzgerald, Ph.D. received her B.S. and M.S. degrees from MIT in 1990 and 1992, respectively, and a Ph.D. from Stanford University in 2000, all in the discipline of aeronautics and astronautics. She has worked at the Jet Propulsion Laboratory, Orbital Sciences Corporation, Sigpro, and Sensant. She founded A.M. Fitzgerald & Associates, a MEMS product development company based in Burlingame, CA, in 2003.

Keith M. Jackson, Ph.D. received his B.S. from Princeton University in 1994 and his Ph.D. from MIT in 2001, both in electrical engineering. He has been active for 20 years in the MEMS and semiconductor industry, and previously worked for Intel Corporation and General Electric.

Charles C. Chung, Ph.D. received his B.A. in physics, astrophysics, and environmental studies from the University of Pennsylvania, M.S. in physics from Duke University, and Ph.D. in electrical engineering from Georgia Institute of Technology. He has worked at Intel Corporation, General MEMS, LV Sensors, Atmel, and Halcyon Molecular.

Carolyn D. White, Ph.D. received her B.S. in mechanical engineering from Brown University in 1998 and M.S. and Ph.D. in mechanical engineering from the University of California, Berkeley in 2002 and 2006, respectively. She previously worked for Lucent Technologies/Bell Labs.