AMENDMENT 2

QUESTIONS AND ANSWERS

Q12. Can you provide a reference defining all the parameters in the example files?

A12. Reference documents cannot be shared publicly at this time. They will be made available at the start of Phase 1. A short description of the files and their parameters is below:

- XRD files: These files contain 2 columns of data, the position (in arcsec) relative to the (004) InP diffraction peak, and the count, which is the intensity of the signal. XRD measurement can be considered to be a Fourier transform in the vertical direction of the electron density of the whole structure.
- PL files: These files contain the peak wavelength in nm as a function of the x and y position on a wafer. The elements in the header are self-explanatory (e.g., date, wafer size, resolution).
- EpiTT files: These files contain the measurements from a 3 wavelength reflectometer and emissivity corrected pyrometer. The important columns are listed below and the other columns are raw measurements used to calculate these important columns.
    - The first unnamed column, which is time in s.
    - The Pyrotemp column, which is the actual wafer temperature in °C.
    - The DetWhite and DetReflec, which are the reflectivity in % at the 2 wavelengths defined in the header.
    - The RLo, RMed, and RHi, the reflectivity at the pyro wavelength defined in the header, for various amplification factors. The Lo amplification factor is most often used.
- Recipe files: This file is the most complex and contains the most amount of parameters. The first part of the file is a series of variable definitions. The actual code to grow a structure is contained in the layer{} section. This code is comprised of steps, which have a defined time, and end with a semi-colon (;). In each step, a series of commands is given. Any command maintains its value if not changed by a step. Important parameters are flows, which are in the form of TMGa_1.source = xxxx. Certain elements have a flow under closed loop control based on a concentration, where the command will be TMGa_1.setconc=xxxx. These flows and concentrations, the temperature, and the structure (combination of layers), are the critical elements to control the PL wavelength.

Q13. Would a solution that relies on a standard software package, e.g. MATLAB be acceptable?

A13. Requiring the purchase of a separate software license would not meet the National Research Council's operational needs. Options using free software (e.g., Octave) may be considered but it is not a preferred solution.

Q14. Could you please elaborate the wording in the essential outcome point 3 - "Demonstrate that PL variations are non-stochastic below ± 5 nm." ?

Is this a question of prediction performance (accuracy) or interpretability?

A14. This is a question of prediction performance. The goal at the end of phase 2 is to obtain a model that can be validated to within +- 5 nm for new data. Phase 1 must demonstrate that this is possible.

Q15. Is there any documentation where we can find information regarding the current models used and the current fabrication process? And is it possible to see an example of the data to be used, especially the MOCVD recipe parameters, PL maps, structure parameters, and X-ray diffraction profiles.

A15. An example of the data is included as a separate attachment accompanying these answers. An example of a model we use to calculate the composition of a single layer is also included. In this model, a target PL and XRD mismatch is given. Composition for a GaInAsP layer is calculated, and based on those compositions two flows are calculated: Ga and As. These flows are the experimental parameters used in the recipe. A good reference for semiconductor parameters and models is the Ioffe institute NSM archives and the references contained within (http://www.ioffe.ru/SVA/NSM/Semicond/).

An example model is also in: Pitts, O. J., W. Benyon, and A. J. SpringThorpe. "Modeling and process control of MOCVD growth of InAlGaAs MQW structures on InP." Journal of crystal growth 393 (2014): 81-84. (https://doi.org/10.1016/j.jcrysgro.2013.10.019)

 To calculate a multi quantum well (MQW) structure PL, the National Research Council uses an in-house model based on the k.p method. A good reference on such models is "Theory of semiconductor superlattice electronic structure", by D.L. Smith and C. Mailhiot (https://doi.org/10.1103/RevModPhys.62.173).

A good reference on the applicable MOCVD process is "Organometallic Vapor-Phase Epitaxy: Theory and Practice", by G.B. Stringfellow (book available through various retail sites).

Q16. What are the details of the current model that is used to calculate PL wavelength?

A16. Empirical data generated through experiments is transformed based on known theoretical models (please see attached spreadsheet for an example), then is fitted using multiple regressions over inputs, both continuous and categorical. The new model should generate predictions that can be validated to within +- 5 nm for new data.

Q17. What is the experimental uncertainty of the PL data and X-ray data?

A17. PL measurement uncertainty is +-2.12 nm. XRD measurement uncertainty is +- 10 arcsec.

Q18. Can this be a web-based application instead of a software application installed in a Windows computer?

A18. No, a web-based application is not compatible with our security and operational environment.

Q19. In Phase 1, given the provided data, if the error rate is higher than 1%, should we proceed with points 2, 3 and 4 in Phase 1, or does the project stop there?

A19. The objective is to automatically read the data with an error rate of less than 1% in order to treat it and build the necessary prediction model. At the end of Phase 1, only proposals that have met this criteria will move on to Phase 2. Please note that, if the original data is missing or erroneous, it is not considered part of the error rate.

Q20. Phase 1: Are the formats of the inputs variable?

A20. The inputs will always be text files (ascii format) but they may contain summary data or all data. The volume of data inside a file is variable and field location within the files may also change.

Q21. Phase 2: Will the newly added data be compatible with the formats of data from Phase 1?

A21. Yes, they are the same inputs.

Q22. Please clarify what you mean by "applicable" in the requirement for Phase 2: "Be applicable to any MOCVD tool of comparable capability"

A22. The National Research Council has multiple MOCVD reactors of the same type of design (i.e., close coupled showerhead). The provided dataset was generated from one of these reactors, but the desired model should be usable with datasets from comparable reactors should the National Research Council acquire further capacity in the future.

Q23. Phase 1 requirements (1): Are the text files digital text files?

A23. They are digital text files in ascii format.

Q24. In the Q&A's it says a small subset of data could be shared as an example of the data format for this project. We would be interested in having access to such a subset, if possible.

A24. This data is included as an attachment to accompany these answers.