Official Transcript of Proceedings

NUCLEAR REGULATORY COMMISSION

Title: Advisory Committee on Reactor Safeguards Digital Instrumentation and Control Systems Subcommittee Meeting

Docket Number: (n/a)

Location: Rockville, Maryland

Date: Tuesday, June 7, 2011

Work Order No.: NRC-933 Pages 1-297

NEAL R. GROSS AND CO., INC.
Court Reporters and Transcribers
1323 Rhode Island Avenue, N.W.
Washington, D.C. 20005
(202) 234-4433

1
2
3
4 DISCLAIMER
5
6
7 UNITED STATES NUCLEAR REGULATORY COMMISSION'S
8 ADVISORY COMMITTEE ON REACTOR SAFEGUARDS
9
10
11 The contents of this transcript of the
12 proceeding of the United States Nuclear Regulatory
13 Commission Advisory Committee on Reactor Safeguards,
14 as reported herein, is a record of the discussions
15 recorded at the meeting.
16
17 This transcript has not been reviewed,
18 corrected, and edited, and it may contain
19 inaccuracies.
20
21
22
23

UNITED STATES OF AMERICA

NUCLEAR REGULATORY COMMISSION

+ + + + +

ADVISORY COMMITTEE ON REACTOR SAFEGUARDS

(ACRS)

+ + + + +

DIGITAL INSTRUMENTATION AND CONTROL SYSTEMS

SUBCOMMITTEE

+ + + + +

BRIEFING ON SOFTWARE PRA

+ + + + +

OPEN SESSION

+ + + + +

TUESDAY, JUNE 7, 2011

+ + + + +

ROCKVILLE, MARYLAND

The Advisory Committee met at the Nuclear Regulatory Commission, Two White Flint North, Room T2B1, 11545 Rockville Pike, at 9:00 a.m., John W. Stetkar, Chairman, presiding.

MEMBERS PRESENT:

JOHN W. STETKAR, Chairman

DENNIS C. BLEY, Member

MICHAEL T. RYAN, Member

JOHN D. SIEBER, Member

ACRS CONSULTANTS PRESENT:

SERGIO GUARRO

MYRON HECHT

NRC STAFF PRESENT:

CHRISTINA ANTONESCU, Designated Federal Official

KEVIN COYNE, RES/DRA

ALAN KURITZKY, RES/DRA

RUSS SYDNOR, RES/DRA

ALSO PRESENT:

TSONG-LUN "LOUIS" CHU, Brookhaven National Laboratory

RICK WEBER, Westinghouse Electric Company*

MENG YUE, Brookhaven National Laboratory

*Participating via telephone

TABLE OF CONTENTS

Opening Remarks/Objectives 4

Overview Digital I&C PRA Research 6

Selection of Quantitative Software 77

Reliable Methods

Bayesian belief network Method 116

Test-Based Method 206

Wrap Up/Future Interaction 302

P-R-O-C-E-E-D-I-N-G-S

9:00 a.m.

CHAIRMAN STETKAR: The meeting will now come to order.

This is a meeting of the Digital Instrumentation and Control Systems Subcommittee. I am John Stetkar, Chairman of the Subcommittee for this meeting.

ACRS Members in attendance are Dennis Bley, Jack Sieber and Mike Ryan. Myron Hecht and Sergio Guarro, our consultants, are also attending this Subcommittee meeting.

Christina Antonescu of the ACRS staff is the Designated Federal Official for this meeting.

The purpose of this meeting is to discuss the ongoing research efforts related to digital I&C probabilistic risk assessment.

During this meeting, the staff will summarize their work on a draft NUREG for development of quantitative software reliability models for digital protection systems of nuclear power plants.

The Subcommittee will gather information, analyze relevant issues and facts and formulate proposed positions and actions, as appropriate for deliberation by the full Committee.

The rules for participation in today's meeting have been announced as part of the meeting notice, previously published in the Federal Register on June 3, 2011.

We have received no written comments. We also have Rick Weber of Westinghouse Electric Company on the bridge phone line, listening to the discussions.

If there is anyone else on the bridge line, Theron, could you open it up and if there is anyone else on the bridge line, could you identify yourselves, so that we know who is out there and listening?

MR. WEBER: Bridge line is open.

CHAIRMAN STETKAR: Thank you. Is there anyone else? Thank you very much.

At this point, we'll -- to preclude interruption of the meeting, the phone line will be placed in the 'listen in' mode during the discussions and presentations to the Committee. Thank you very much.

A transcript of the meeting is being kept and will be made available, as stated in the Federal Register notice. Therefore, we request that participants in this meeting use the microphones

located throughout the meeting room, when addressing the Subcommittee.

The participants should first identify themselves and speak with sufficient clarity and volume, so that they may be readily heard.

We will now proceed with the meeting, and I call upon Mr. Alan Kuritzky of the Division of Risk Analysis in the Office of Nuclear Regulatory Research, to provide introductory remarks.

MR. KURITZKY: Thank you. As Dr. Stetkar mentioned, I am Alan Kuritzky with the PRA Branch in the Division of Risk Analysis in the Office of Research.

We me today are Louis Chu and Meng Yue from Brookhaven National Laboratory, who are some of the main investigators doing this research up at BNL.

There are some additional people that have been working with them, primarily Gerado Martinez-Guridi, who is not here today, has done some of the work, particularly in some of the previous studies, as well as some of the work that we're going to discuss today, and the purpose of today's briefing is really just to give a status update to the Subcommittee, as where we stand right now with our Digital I&C PRA Research.

We briefed this Subcommittee pretty much almost exactly a year ago. It was last June, was the last time that we were here, and so, we have accomplished a number of things in that time period, and we want to bring you up to speed on that, and we're also very interested to get any and all feedback from the Subcommittee on the software PRA activities that we're currently pursuing.

Since this is primarily a meeting to let you know what we're doing and get your feedback on it, we are not requesting a letter at this time.

The next couple of slides are ones that I think you've seen at multiple briefings before, haven't really changed, the background and objectives.

As we all are aware right now, the process for licensing digital systems at U.S. nuclear power plants is purely on a deterministic basis. As an example, the Oconee TELEPREM, the RPS and ESFAS that was approved about a year and a half ago was done purely on a deterministic basis. There was no risk arguments provided with that submittal. The NRC is fully capable of approving digital systems on that basis and can continue to do so.

However, as consistent with the Commissions's 1995 PRA Policy Statement, we want to

try and incorporate PRA into any of the regulatory decisions, wherever we can, and so, we're trying to pursue that for digital I&C systems.

The concern is that, you know, the risk-informed analysis process for digital I&C systems, right now, is not mature enough for broad regulatory use.

We're trying to make headway in that regard, and that's the impetus for the work that we're going to describe for you today.

CHAIRMAN STETKAR: Alan, are you -- you mentioned the Oconee upgrade was evaluated, based on deterministic criteria.

Have you or NRR received any indications that applicants may be interested in using a risk-informed basis for upgrades to existing systems, at any sites?

MR. KURITZKY: I can't really speak for NRR, although I probably would have heard something about it, I assume, if something had come in.

All I know is that a few years ago when the Digital I&C Steering Committee was stood up, I think in early 2007, after the industry came and met with the Commission in October 2006.

They set up the Steering Committee. A

number of task Work Groups were established in different areas, in the digital I&C to work with industry, to resolve outstanding issues, to allow digital instrumentation to move forward at nuclear power plants.

One of those Working Groups Number 3 dealt with PRA, and one of the main focuses of that Working Group from the industry side was to try and use a risk argument to justify that you do not need to have an automated diverse actuation system.

So, they made a risk -- they submitted, I think, an actual -- I guess it would be an NEI White Paper, I think, to try and justify, you know, why it was not -- you didn't -- we didn't actually -- there would be a risk increase by having the system.

I'm not going to go into those details, but the staff did not accept that reasoning, responding in a letter, saying that we did not, you know, buy into that, and that's pretty much the last that I've heard of risk.

CHAIRMAN STETKAR: Okay, the second question, the part B of the question, and maybe this is a good lead-in, as far as, I'm aware of one current applicant for Design Certification and a COL, in particular, COL applicant.

We have not seen any details of the application, so this is also somewhat hearsay, but I think fairly good hearsay, who is interested, at least, in using risk information to justify technical specifications for a new plant design. It would be through the COLA applicant, not through the certified design, but obviously, they're inter-related.

I have no idea how that initiative may relate to digital I&C systems, because this design, as do all of them, has, you know, fully integrated digital instrumentation and control.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: So, I was curious as to whether you were aware of that, also, because that may be a forcing function in some sense, for this research work, because as I've said, we have not gotten into that review far enough to really understand the scope of what they're going to do, in terms of the breadth and the depth of the risk information, or how they might be planning to, you know, adjust technical specifications, in particular with the I&C system.

So, I really don't know where that may lead, but that seems to be, at least in the licensing arena, one initiative that's out there, if there is -- if there isn't anything on the upgrades of existing

plants.

MR. KURITZKY: Right, and I think, you know, it's interesting, I haven't heard of that.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: And it's interesting to hear, because there has been some concern on our part, as to when we're going to have industry really engaged on this aspect, and I guess --

CHAIRMAN STETKAR: Well, this is not trying to justify, you know, automated or manual diverse actuation systems.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: This is literally a risk-informed tech spec initiative, going in at the COL stage.

We're learning more about that. That's about all I know, so, I was curious, whether you had heard much about that --

MR. KURITZKY: No, but I --

CHAIRMAN STETKAR: -- because as I said, it's -- it may be the nearest term point where we actually run into issues regarding models for digital I&C systems in a PRA, you know, quality of the data that they use and things like that --

MR. KURITZKY: Sure.

CHAIRMAN STETKAR: -- in a, you know, real world application.

MR. KURITZKY: Right, it will be interesting to see how that plays out.

I mean, the one thing I do know is that back under Task Working Group 3, under the Steering Committee, we did put together an ISG-03, which dealt with how do you include digital systems in the PRA model, but that was specifically for these -- you know, design cert PRAs in COL, and it was kind of based a lot on what was accepted for the AP600 previously, and we tried to draw a tight balance on the applicability of that.

And so, I don't know whether something like this might fall under that or not.

CHAIRMAN STETKAR: This attains -- in general, this would raise the bar on both level of detail and quality of the models, compared to what, at least the staff has accepted in previous design certification PRAs, you know, strictly for the purposes of the design certification process.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: Because is now, actually, an application.

MR. KURITZKY: Right, yes, I have --

CHAIRMAN STETKAR: It's a little bit different.

MR. KURITZKY: Well, I haven't heard about it, but I'm interested to see --

CHAIRMAN STETKAR: You might want to follow up on that.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: We can talk about that offline a little bit later.

MR. KURITZKY: Yes, right.

CHAIRMAN STETKAR: Continue, sorry.

MR. KURITZKY: No problem. Okay, so, the objective of this work, again, based on some of the discussion we just had, is to identify and improve and if necessary, develop methods, tools and guidance, so that we can include digital systems into plant PRAs, as well as incorporating them into the NRC's risk-informed initiatives, like the risk-informed tech spec that Dr. Stetkar just mentioned, reactor oversight program, you know, maintenance role, those are all -- you know, any number of programs and processes that the NRC has, that have risk-informed nature, and until we have modeling methods and tools for digital systems, they're always going to be kind of an outlier of that process, and not really seamlessly fit into

it.

Okay, the research in this digital I&C area has been going on since back in the early 2000s. In that time period, of course, you've accomplished quite a bit.

There has been some work that we've done on failure mode identification and analysis, BNL, looked in detail at a digital feedwater control system and did a very detailed FMEA of the hardware of that system. There is -- and that's documented in the NUREGs at 6962 and 6997.

Of course, there is other work in that area that's been done by other areas with the NRC, as well as outside the NRC. The Digital I&C Branch in the Division of Engineering in the Office of Research has done a lot work in the digital I&C area, including failure mode identification and analysis, that's Russ Sydnor's branch.

There is work that's been done by EPRI and others outside the NRC, as well as internationally.

There is some -- there is an activity going on right now with -- under the Organization for Economic Cooperation and Development, OECD, Nuclear Energy Agency, NEA, Committee for the Safeguards of Nuclear Installation, CSNI and Working Groups there.

They have a number of Working Groups, one involving risk, and one of the activities that that Working Group Risk is current engaged in is trying to come up with failure mode taxonomies for hardware and software for use in the PRAs, and I'm going to talk to you for a few minutes about that, and the end of this -- a couple slides at the end of the presentation, because we just completed the first meeting for that activity just a few weeks ago.

And so, certainly, a lot of activity that's addressing the issue of failure mode identification. It's still a big issue. There is still the idea of completeness of failure mode identification with digital systems, you know, being a much more complex undertaking than the typical mechanical or fluid system in the PRA -- current PRAs.

So, that's one area that some work has been done. We've done a lot of work in reliability modeling. There is a couple of studies, but we looked at actually, a spectrum of methods that we applied to the digital fuel rod control system a few years back.

Ohio State University and some of its subcontractors applied dynamic reliability modeling methods to the system. They applied two different methods, a dynamic flow graph method and -- which Dr.

Guarro is intimately familiar with, and also a Markoff modeling cell-to-cell mapping technique approach.

I think, you know, just to clarify, again, when we talk about dynamic modeling methods in this context, dynamic means that the modeling -- the model explicitly attempts to account for the interactions between the control system being modeled and the processes being controlled at the plant, and the timing of those interactions.

So, in the traditional methods, essentially, are other methods that are commonly used out there right now for PRAs, and don't try to account for those interactions explicitly.

Now, one of the things that we did learn from the -- in the OSU work was that there is, you know, some concerns, or some doubts were raised about the usability and the computational practicality of applying those methods, and obviously, it's a more complex undertaking.

You definitely get a more accurate evaluation of the system. The question will always be -- well, may not always be, but the question is certainly right now, do you need that level of accuracy to give you an adequate model for the PRA?

CHAIRMAN STETKAR: Unfortunately, I'm not

intimately familiar with the Ohio State work. I've read a couple of reports on it, but not very much in detail.

How does your characterization of that dynamic modeling approach from Ohio State differ from what I read that this current effort is doing, in terms of integrating thermal hydraulic models with Markoff models of the software and hardware integrated system? Isn't it the same?

MR. KURITZKY: Well, I'm probably going to have to --

CHAIRMAN STETKAR: You know, you're characterizing your current modeling methods as being traditional methods and the dynamic methods as being some different approach to modeling.

But it seems to me that you're approaching something that at least on the surface, not knowing much of the details of the Ohio State methods and models, it seems to be very similar.

MR. KURITZKY: Okay, there definitely is a sense of convergence there, in some regards.

I think one thing to kind of -- to lay out first, is that the traditional methods, they also are not -- I mean, what BNL ended up doing wasn't so traditional, also.

CHAIRMAN STETKAR: Right.

MR. KURITZKY: Just because the nature of the system was such that you couldn't do the normal fault tree modeling, and so, they ended up having to come up with a simulation tool, which was based on the source code of the system and it generated failure paths for the system.

But in terms of the work that we've been doing with BNL, the hardware and the software are two distinct entities, and so, the hardware modeling did not address this dynamic interaction.

Now, if the software it becomes a little more difficult to step aside from that, because if you want to -- you can do it -- there is a lot of approaches out there, as we may discuss coming up, on software -- ways to quantify software failure probability.

Some of them, or a lot of them, don't really account for the context of the software, and the context of the software is something that we feel is important to account for.

There is the approach actually from ASCA that does a very -- it was very focused on looking at the context of the software. The only focus on mostly on the failures, because those are the ones that are

1 believed to be the biggest contributors to risk.
2 And so, the work that you were just
3 referring to, that we're doing in the software area,
4 tries to kind of mimic that type of approach, where we
5 -- we use risk information to come up with a different
6 context for the software and then use risk information
7 to try and reduce the test burden.
8 But for the software, we're kind of -- at
9 least in our mind, you need to include that. There is
10 plenty of approaches that don't deal with that, so you
11 don't have to go that route. But we kind of felt it
12 was important.
13 CHAIRMAN STETKAR: Thanks, and I know
14 we'll hear, you know, more details about the path
15 you're headed down, later in the meeting.
16 I was just curious about, you know, is the
17 distinction between those two sub-bullets on your
18 slide as clear as you might present it --
19 MR. KURITZKY: Right.
20 CHAIRMAN STETKAR: -- or are they —- are
21 we morphing into something that's --
22 MR. KURITZKY: Well, that's it, this is
23 kind of like a snapshot in time.
24 CHAIRMAN STETKAR: Okay.
25 MR. KURITZKY: And most like the

traditional method.

CHAIRMAN STETKAR: All right.

MR. KURITZKY: This bullet is from when we were doing the original studies, where we only focus on the hardware.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: Whereas, to be clear, the Ohio State work did incorporate software, to some extent. There were some aspects of software that were not accounted for, but software was accounted, to some extent, and I think that's one of the reasons why it was more of a natural fit with the dynamic modeling, because it was accounting for the software in the operation, et cetera.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: Whereas, the BNL approach did not delve into the software, and so, it didn't have a need to --

CHAIRMAN STETKAR: The initial -- yes, okay.

MR. KURITZKY: Initial -- the initial model, yes, right.

CHAIRMAN STETKAR: Okay, thanks.

DR. HECHT: May I ask a question, with respect to what we're talking about, the kinds of

digital systems that we're talking about.

If we're talking about something like an RPS, which basically works on pretty much a one-shot basis, we're saying that when you reach a certain condition, do something, as opposed to a feedwater control system or something, where you are talking about closed-loop interaction constantly.

What are we talking about in the PRA? Are we talking about both?

MR. KURITZKY: Okay, I'm glad you brought that comment actually up, because that was one of the points I wanted to make from the previous slide, when I was talking about the dynamic approach.

Besides the fact that there was some concerns raised about the computational practicality of it and the complexity of the dynamic modeling, the other aspect was, how important is it, particularly if we're looking at, you know, stand by protection systems, like an RPS or a safety injection actuation system.

For a control system like the fuel rod control system, you can envision that it might be much more important to have it. But for a stand by protection system, which is just going to sit there and read values, and it's going to try and compare

them to some trip set point, and once it hits that point, it's going to send out a signal to actuate, a lot of that -- you know, we don't believe it's necessary to have that level of complexity.

So, yes, that's a very good point. We're primarily talking about protection systems.

There are some control systems that are in a PRA, for various -- you know, at various points. Primarily, they're there as an initiating events, in which case, we don't really need to model them in detail because a lot of that can be used with just data on the system failure, you know, spurious initiation of the system, or failure of the system.

But as far as mitigating systems in the PRA, it's primarily the protection systems, which are essentially in a stand by mode.

Now, again, it depends, that's based on what's currently out there being proposed, like for instance, the TELEPREM, which is really an RPS and a safety feature actuation system.

In the future, you can envision, you know, service water systems, which are on a normal operating mode, or whatever, can -- you know, can all be digitally controlled, and other types of systems might get digital controls, which might blur those lines a

little bit.

But I think for right now, we're primarily looking at the protection systems, and that's one of the reasons why for the time being, we feel comfortable looking at, you know, not incorporating the dynamic aspects right now.

But it's something that I think we can't totally shut the door on because it may be important for certain systems, and those systems may become more prevalent.

CHAIRMAN STETKAR: I was going to say, I think we need to be aware of new plant designs and the features of their digital I&C systems, even within the traditional protection function, if you want to characterize that way, because for example, I know of at least one design that does incorporate a type of control, ongoing active control, in their protection system, and particularly, it's a secondary -- it's a controlled programmed secondary cool-down rate to achieve high pressure injection.

That's at least one that I know of, which is not simply a protection, you know, as Myron characterized it, a one-shot.

This is something that's invoked under certain conditions, at which point a control system

takes over for a cool-down rate to achieve a certain desired condition, by a certain desired time.

MR. KURITZKY: Yes.

CHAIRMAN STETKAR: And I'd hate to see the research being focused solely with the notion that the only type of protection systems that we're ever going to see are these, you know, go/no-go bimodal type decisions --

DR. GUARRO: Yes, because of the --

CHAIRMAN STETKAR: -- because we know of at least one system that's out there, in a design that's currently under certification, that involves some measure of this type of control function.

DR. GUARRO: Yes, if I can add to that, and Alan, you know, I think you know already, my thinking on this, but I would reinforce this suggestion, because one, unless, you know, there is a -- I would say, a very firm rule by which the functionality of software is forbidden from evolving in the future, which is very difficult to do, okay, so, software is the tempting apple in the Garden of Eden, okay.

You can put as much complication in it as you want, and the temptation to do so, that's very strong. So, I can guarantee you that more

functionality will creep into even the safety part of the thing.

But even aside from that, you know, we have to remember in a risk perspective, you know, the trouble that nuclear plants have had in the past, were not because the RPS didn't work.

It was because there were interactions that happened between the RPS and human functions and so forth.

So, I think to keep the mind open to that is important, because you know, the part -- part of the risk that we control well is perhaps, that confined to the safety system. It's the other part that sometimes, creeps in, in ways that are hard to anticipate. So, that's the comment.

CHAIRMAN STETKAR: You know, I think part of the message is that I'd hate to see the research program progress to a point where you need to find a very square, very black box that you're operating within, spend a lot of effort to develop models, methods and data that fit within that box, and then discover that that box doesn't work very well for particular types of integrated protection control systems that we may see in new reactor designs, and discover that, you know, you then need to go back and

redefine the box, and start out again.

It strikes me that if you accept the challenge of trying to do the integrated system, the simpler sort of bimodal -- you know, the evaluation of the bimodal functions may fall out as a, I would imagine, a rather simple consequence of that.

I don't know that to be true, because I don't know how software works, but --

MEMBER BLEY: I have a slightly different question. Back to the Ohio State work and your -- the computational complexity issue you talked about.

That's a broad term and I can think of many things you might be worried about from there. We could be worried that the models are so complex, we might not believe them, might not have a way to verify them.

It might be that the human input to make these models work is a large and expensive effort. It might be that the codes they were using were -- took a long time to run, but I'd wonder if that had to do with them being further type kind of things in the computers they were using, whenever this -- this was few years back, as I recall.

So, can you talk a little bit about what the problems were, from your point of view, with this

computational complexity issue?

MR. KURITZKY: Unfortunately, Dr. Bley, I can give you a simple answer to that, and that's no.

MEMBER BLEY: Why is that?

MR. KURITZKY: Okay, so, let me -- I will mention that the issue, as far as being a prototype and how -- and certainly, that's an influencing factor and as we become more familiar with it, that should go down.

Just the power of computing, just from a few years ago, and in a few years in the future, will go up by so much, that some issues along those lines will tend to minimize.

As far as some of the first items that you mentioned, which wouldn't tend to necessarily disappear with time, I can't really speak to them.

That work -- we used it for the -- again, like you mentioned, this was a few years ago, and for the first couple of briefings after that study was completed, we used to bring along someone from the group that did that work, to be able to answer such questions.

But since it's longer -- three years back or so, we've stopped bringing that person along. I don't know, you know, Russ, is there -- do you want to

speak to that, at all? I mean, do any of you, probably the perfect person --

MEMBER BLEY: I guess the real thing is, my concern is, are those issues, issues of the sort that would affect our confidence in using them? Are they things that might improve over time, just basically, what is our problem with that approach, and I understand it could be a range of things.

So, if somewhere along the line, somebody could get back to us about that, it would be better than, you know, just throwing something out there.

MR. KURITZKY: Right.

MEMBER BLEY: But you might have an answer now, that's all. I didn't want to cut you off.

MR. SYDNOR: Is this on?

CHAIRMAN STETKAR: It is, just identify yourself.

MR. SYDNOR: My name is Russ Sydnor. I'm the Branch Chief of the Digital I&C Branch in the Office of Research, and we were the sponsor of this work with Ohio State, and I would categorize our decision more as a combination of timing and comments that ACRS actually provided to the DRA, based on the Brookhaven work, and the fact that --

MEMBER BLEY: Help me out, DRA is?

MR. SYDNOR: Division of Risk Analysis.

MEMBER BLEY: Okay, thank you.

MR. SYDNOR: Alan's group. Alan was -- about the same time we were thinking about bringing in Ohio State work for ACRS review, the BNL work was reviewed and we felt a lot of the comments that ACRS had at the time on that work, pertained to this work, too, and that was -- you know, are you -- how are you addressing software, was probably the biggest comment.

And so, Alan is not -- hasn't finished getting through his story yet, but one of the reasons we're pursuing what Alan is pursuing, and he's going to talk about today, is to try to answer that question.

We did not feel that the methods were not viable or not -- could not be improved or it was more a question of spending additional resources, to do a second benchmark or model study, when you still had this major issue over shadowing --

MEMBER BLEY: Fair enough.

MR. SYDNOR: -- the work. So, that was really the primary.

So, a lot of your questions about, you know, could the methods be improved over time? I think we thought that could happen, but we weren't

prepared to invest more resources in doing that, until we had addressed this other issues.

MEMBER BLEY: Right.

MR. SYDNOR: And that's really the primary purpose of the discussion today, is to talk about how we're addressing software.

MEMBER BLEY: Okay.

MR. SYDNOR: Or how we might address it.

MR. KURITZKY: And I think also, just to follow on with that, it gets back to what Dr. Stetkar was talking about, as far as, you know, limiting the research to a box, and we're pursuing certain approaches through BNL.

There are other approaches that are also being pursued in other parts of research on some things, and there is a whole big picture that's going to have to be synthesized together at the end.

We're not really closing any doors, as Russ was just mentioning. You know, we're focusing on different things at different times.

We have IOUs for a lot of stuff that are floating out there, and one of them is determining whether or not the need for dynamic modeling is -- you know, how important is that, and it gets to the issue as to, do we need it for protection systems?

Are there types of system designs that, even in protection systems, where you do have a dynamic interaction occurring?

As Dr. Guarro mentioned, there is going to be -- certainly, the wave of the future is to make the leadership do more and more and more, no matter how much -- we can't essentially get them to do what we want, right now.

So, as future capabilities increase, it's going to be hard to --

CHAIRMAN STETKAR: Yes, it's really difficult to predict the future, but as I said, at least, you know, make sure that the approaches that are being evaluated are responsive to the scope of designs that we're currently aware of.

MR. KURITZKY: Right, and I think to get --

CHAIRMAN STETKAR: But looking forward to, you know, designs that are in the certification stage, you know, again, looking past that is something that starts to get at a lot of crystal balls.

MR. KURITZKY: Right, it's murky.

CHAIRMAN STETKAR: We can't really do that.

MR. KURITZKY: But even at this stage,

there is a tradeoff here between trying to essentially take a bit that we can chew, and swatting at something we just can't get our hands around.

CHAIRMAN STETKAR: Well, you know, we learned, I think painfully in some cases, back in the early 1980s, that perhaps, trying to scope the problem down to something that's so simple, that somebody can try to do, might not be the most productive way to tackle the problem, because you then quickly learn that your simplified methods and approaches don't work very well, when you start to integrate them into, you know, larger models and things like that.

So, I understand, you're right, Alan, it's a balance between resources and schedules and things like that. It's just, from a purely research perspective, you know, being aware of the potential dangers of simplifying the problem that you're currently working on, to such a well defined little box, that it may constrain you too much.

MR. KURITZKY: And that's something that I -- and that's a question that we ask ourselves, all the time, it's something that we -- is certainly in the forefront of our minds, it's, how well is the work that we're doing scale up --

CHAIRMAN STETKAR: Yes.

MR. KURITZKY: -- to a broader application --

CHAIRMAN STETKAR: Yes.

MR. KURITZKY: -- and it's something that we're certainly cognizant of and keeping in mind.

But there is also the recognition that some of these more complex applications will require such extensive modeling that it may be -- for those cases, it may be that you could get by with a good enough model, without quite that level of complication.

But more importantly is that we may feel that a more simpler approach may work for a number of different systems or applications, and a more complex one may be needed for others.

So, right now, we're trying to see what can we get by with, because we don't want to have to do a very big complex thing, for everything --

CHAIRMAN STETKAR: Right.

MR. KURITZKY: -- if there is a lot of it that we can do with a more simpler approach, and so, that's kind of the attack that we're taking, but we're certainly conscious that what we do is not necessarily going to be the end-all and be-all of digital system modeling.

We're trying to get a foot-hold in and then be able to build off of that. So, you know, your points are well taken.

CHAIRMAN STETKAR: I'll let you get to two of two.

MR. KURITZKY: Okay, so moving -- yes, I think we -- we pretty much hit most of that, on that slide.

So, other work that's occurred over the previous years, we've looked at some operating experience. BNL has studied some methods and data from non-nuclear industries, primarily defense aerospace telecommunications.

They collected some hardware failure data and developed the database of some of the basic hardware, digital I&C hardware components. That was documented in an internal letter. It showed up in NUREG/CR-6962, and if your call from that NUREG/CR, or from the -- what we briefed the Subcommittee on a few years ago, it was pieced together from a number of different varied data sources that were, in fact, so varied that the uncertainly balance on some of these component failure modes were so great that it really was, you know -- could be considered limited usefulness, and also, it necessarily didn't apply to

nuclear plants, anyway.

But as just a basic starting point of something to use as a surrogate, to try and test out methods, it was something that we had put together.

BNL also looked at software experience -- experienced both nuclear and non-nuclear, and that write-up was included in an old appendix to NUREG/CR-6962 in draft.

It was pulled out before that NUREG was published. It's now going to be reincarnated as a BNL technical report that's going to be publicly released some time later this year.

MEMBER BLEY: Essentially, in the form it was when we saw it several years ago? But has it been expanded, some?

MR. KURITZKY: A little bit has been adapted because we've learned stuff since then. Some stuff, actually, has been taken out because there were certain aspects of it that we felt no longer applied, and so, we took them out.

The part about BNL's idea of how to put together failure modes and failure cause taxonomy on software, that is included in there. The review of all the data is included. The detailed review of certain events and how they're going to apply to that

-- to the proposed taxonomy that BNL came up with, that will all be included.

But there was, I think, some stuff in there on -- commenting on other types of methods, which we took out, because that was essentially covered by other reports that had been released since then.

But the bulk of the software experience stuff and software analysis stuff is still going to be in there.

CHAIRMAN STETKAR: So, that's going to be just issued as a BNL technical report --

MR. KURITZKY: It will be issued --

CHAIRMAN STETKAR: -- and not another NUREG --

MR. KURITZKY: -- as BNL report, but it will be publicly available, not a NUREG.

CHAIRMAN STETKAR: Any idea on --

MR. KURITZKY: Time line?

CHAIRMAN STETKAR: Yes.

MR. KURITZKY: Depends how soon I can -- I kicked the football onto Louis' court a few months ago, but he's been busy doing this, since then. So, hopefully maybe some time in the Summer, I don't know. I don't know if you have a schedule for that.

CHAIRMAN STETKAR: I mean, I think we'd be really interested in seeing that. I'm just curious about --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- what the time is.

MR. KURITZKY: Yes, unfortunately, it doesn't have a -- it's the one thing that's not working to a dead line, so, it's the one that keeps getting punted every time something else needs to get done.

CHAIRMAN STETKAR: Understand.

MR. KURITZKY: But hopefully, some time in maybe the Summer.

CHAIRMAN STETKAR: Okay, we'll just -- we'll keep track of it.

MR. KURITZKY: Okay.

CHAIRMAN STETKAR: We'll leave that with you.

MR. KURITZKY: So, once we get it out, we'll have --

CHAIRMAN STETKAR: Yes, okay, good, good.

MR. KURITZKY: All right.

CHAIRMAN STETKAR: And even if you get -- well, we'll talk about later schedules.

MEMBER BLEY: Would I be putting words in

your mouth, and appropriately, if I assumed the reason it was not coming out as a NUREG is, you think it's more of historical significance and it's not up to the -- where your work might be currently, or where you want to publish something in this area?

MR. KURITZKY: I think that's a good assessment. It's not really the -- where the push is, right now. It's kind of more tying up the loose ends from before, because it got dropped from that one NUREG. So, you just want to get it out in the public domain.

MEMBER SIEBER: I'd sure like to ask a question that's a little bit off the subject, and probably not too wise.

Has anybody done a study that compares digital systems to analog systems, as far as failure rates are concerned?

MR. KURITZKY: Not to my knowledge, and you know, that's interesting, because as you'll see, our approach that we're going to propose for doing this software thing kind of relies on that fact that it may be difficult to demonstrate an exact reliability for the software, failure probability for the software, because it requires such extensive testing, but that there might be some acceptable

limit, if we could just show it's at least as good as 'x', that may be good enough.

And so, for 'x', we're kind of using -- saying, "Hey, what's the existing reliability for the analog systems," and let's accept that as the 'x'. We want to be at least as reliable as those.

But the reality is, for a lot of the systems that are analog right now in PRAs, they're not modeled -- like, the rad protection system in most PRAs doesn't have a very detailed modeling, and it's a very --

MEMBER SIEBER: Yes, but it's pretty simple, and I don't think that you can do, from an analog standpoint, all the things you can do from a digital standpoint, as far as curve-fitting and choice making.

On the other hand, right now, we really don't have a good handle on reliability of digital systems compared to analog systems, upon which all the current PRAs are based.

To me, it's sort of a missing piece, and I haven't decided in my own mind, whether there is a reasonable relationship between analog and digital, from a complexity and capability standpoint, but I think from an overall characterization of the safety

of the plant, that would be a factor.

MR. KURITZKY: Yes, I think -- well, just a little information on the modeling analog systems.

Idaho National Lab did some very detailed studies on the various reactor protection systems --

MEMBER SIEBER: Right.

MR. KURITZKY: -- with the different vendors, and a number of PRAs have actually included detailed models of reactor protection system, the majority don't.

But the reality is, even though detailed studies, that ultimately falls on eventually two basic things, which is the common cause failure of the reactor trip breakers to open, or some just live -- some failure values that is set for like the logic, and they're really -- there is not much that is, I would say, you know, high quality modeling and quantification there.

It's really more of a judgment thing, so, we can make a comparison based on those values, but how much faith you put in those values, you know --

MEMBER SIEBER: Well, you know some of the things are common between analog and digital. For example, sensors and sensors fail, in my old experience, more frequently than any of the other

components, and contact failures and so forth.

So, somehow, you need to know what the pieces are, in order to make an improvement, you know. If reliability doesn't meet your goal, you need to find out what parts of it don't meet and what you can do about it.

Anyway, that's sort of a distraction from what you're talking about here, and I apologize for that, but I am still curious.

MR. CHU: Louis Chu, BNL. Regarding your question, in the past, I've seen the paper published by the Korean -- Korea Atomic Energy Research Institute, where they compare --

MEMBER SIEBER: Oh, really?

MR. CHU: -- analog with digital design. The results are somewhat surprising. They seem to say that digital system has higher failure probability. I don't know the detail of their modeling.

But it would be interesting to look into the detailed modeling --

MEMBER SIEBER: Now, is that paper available? Could you give me a reference to that?

MR. CHU: Yes.

MEMBER SIEBER: Okay.

MR. CHU: I should be able to find that.

MEMBER SIEBER: Okay, I can give you my email address.

MR. CHU: Okay.

DR. HECHT: John, when you asked your -- John, when you asked your question --

MEMBER SIEBER: Yes.

DR. HECHT: -- were you thinking about current -- I mean, we have to compare what kinds of systems we're talking about.

When we say digital I&C systems as opposed to analog I&C systems --

MEMBER SIEBER: Right.

DR. HECHT: -- first of all, are they the control side or the safety side, and secondly, what generation of electronics are we talking about, because the underlying electronics, of course, is much more reliable now, than it used to be.

MEMBER SIEBER: Yes, and for example, there were -- if you go back 50 years, they used mag amps for things, which were notoriously inaccurate because they're temperature sensitive.

DR. HECHT: Yes.

MEMBER SIEBER: But I'm referring, in my question, or what I'm thinking about is protection systems, as opposed to control systems.

DR. HECHT: But even there, older generation systems --

MEMBER SIEBER: There is improvement, yes.

DR. HECHT: Yes.

CHAIRMAN STETKAR: Can we -- we're going to run into a real time problem here.

This is kind of interesting discussion for a historical perspective, but you need to let Alan finish the second half of two of two, and see if we can get onto Brookhaven's discussion.

But the way, Louis, if you could, if you do track down that paper, could you make sure you get it to Christina, so that we can distribute it to the --

MR. CHU: Sure.

CHAIRMAN STETKAR: -- you know, full Subcommittee.

MEMBER BLEY: You didn't refer to it in here, did you? You had some Korean papers in here, okay.

MR. KURITZKY: Yes, but I don't think that was one.

MEMBER BLEY: Okay.

CHAIRMAN STETKAR: Anyway.

MR. KURITZKY: We'll get it.

CHAIRMAN STETKAR: Thanks.

MR. KURITZKY: Okay, lastly, just to wrap up two of two.

Oak Ridge has also done some work on looking at operating experience. Primarily, that effort was trying to come up with looking at the data to see if there was enough information on the digital system failures that you could put together some kind of taxonomy of failure modes and failure causes, and the outcome of that work, I think, was pretty much that, no, there is not sufficient detail in the data, to be able to accomplish that function.

The best database they looked at was the EPIX data, which is run by INPO, but even that, in many of the cases, did not get to the level of detail that you need to really break down failure modes and failure causes.

In fact, I think in a lot of the advances, it's even difficult to tell whether they're talking about an analog system or a digital system, because some of the vendors are the same and stuff.

MEMBER BLEY: Just to put it in perspective, when was the Oak Ridge work done?

MR. KURITZKY: That work done? That was done under Russ's group, but I think that report

actually came out just about a year ago, or so. The work was probably done two or three years.

MEMBER BLEY: So, it's reasonably recent?

MR. KURITZKY: Yes.

CHAIRMAN STETKAR: They also looked -- I mean, they went out and they -- if it's the report that I recall, they looked at NASA and other industries, also, didn't they or not? Is it a different report?

MR. KURITZKY: Russ, do you want to --

MR. SYDNOR: This is Russ Sydnor, again. Yes, that -- directly, two efforts, there.

Primarily, it started out as an effort where we asked Oak Ridge to look at operational experience, primarily non-nuclear, to -- and it was, again, from apparently ACRS interactions going back to 2006 or 2007, where there was questions that -- you know, look at how other industries are doing digital.

And so, they tried to identify a number of failure databases that might be available and they identified about a dozen different potentials, some of which you could get access to, some of which you had to buy, and things like that.

So, eventually, we gave them enough resources and they were able to explore, and we're

really looking for is, you know, did any of those databases actually give us enough insight that we could actually develop a taxonomy and failure modes and things like that.

Unfortunately, that was not the case, and so, it documents the effort, but it didn't really lead us to, you know, a conclusion, a useable conclusion.

CHAIRMAN STETKAR: Well, it's a useable conclusion. The answer was no.

MR. KURITZKY: Okay, so that moves us onto our current activities, and what we're focused on right now, of course, is trying to incorporate software failure into the models, and the term software failure itself is somewhat contentious.

There are many people who don't believe that software fails, per se, that it just -- it does what it's programmed to do, good or bad, and but for the sake of a discussion and ease of terms, we've heard of software failure.

Clearly, there are a lot of software based systems out there that have failed in part or entirely due to software problems.

I mean, there is planes that crashed. There are rockets that explode. There are grids and telecommunication networks that go down. There are

medical devices that have lethal effects on people, and all because -- or at least in part, due to software issues.

So, for the sake of convenience, we refer to it as software failure.

One of the things that we looked at first was based on a comment we received at one of the Subcommittee briefings was, hey, let's, you know, make sure that on a philosophical basis that it makes sense to model software failure, you know, that software -- you can model it probabilistically to include into a PRA model.

So, a couple of years ago at BNL, we pulled together a number of international experts in software reliability. We put the concept to them, to come up with the philosophical basis, and they unanimously agreed that yes, it makes sense to model software failures probabilistically, when it came up.

They established that basis. We documented it in a BNL technical report that's publicly available. We discussed that previously, in the previous briefing.

The conclusions of that panel were very consistent with those that came out of the 1997 National Research Council study on safety and

reliability issues, the studies of digital systems at nuclear plants.

So, with that under out belt, we've moved forward into looking at the various methods for quantifying software reliability.

BNL did a survey of those methods. They looked at stuff that had been previously -- research that was sponsored by the NRC. They looked at previous research sponsored by NASA. NASA has been a big leader in this area. They have a lot of core software involved in their systems and their missions, and they've done a lot of work in this area.

So, we availed ourselves of all of the work they had done. We also looked at what was being done internationally. We also -- BNL did a literature review, looking at books and papers on software reliability and going through proceedings of recent conferences dealing with software reliability and with PRA, and from all those various methods that were identified, they went and reviewed them and also came up with a set of desirable characteristics they felt that a method should have, if it's going to be used in the PRA.

Things like, it should account for the quality of the software life cycle activities. It

should be able to account for any testing results or operating experience associated with the software, and the previous report that's -- that BNL put out documents the search, what they thought about those various methods, and that leads us to the current report that is out now for public comment, the draft NUREG, which now says, "Hey, let's compare all those methods to the criteria. Let's come up with, you know, one or two methods that we can now go test out in a case study and approved concept study, to see how well it will work," and that is where we are right now.

We've identified two methods, the Bayesian belief net method and statistical testing method that we want to further pursue in case studies, and that's what Louis and Meng are going to talk to you about today.

We're going to describe our approach for picking those methods. We're going to describe our approach that we're going to undertake to test out those methods and basically, what we're going to do is, because all of the methods have some kind of limitations, we are going to use the Bayesian belief net method to come up with a prior distribution of the failure probability of the software, based on software

life cycle qualities, et cetera.

And then, use Bayesian inference to update that distribution with statistical testing results, and again, that should be the topic of what Louis and Meng will talk to you about.

MEMBER BLEY: Alan?

MR. KURITZKY: Yes.

MEMBER BLEY: I've got one thing that's been nagging at me since I read through your report.

MR. KURITZKY: Only one?

MEMBER BLEY: One that's really nagging at me.

MR. KURITZKY: All right.

MEMBER BLEY: And your last two slides, you went through, ghee, we've tried to look at failure modes and we just can't find anything, and now, we're going off and testing methods to see how well they do, to give us failure rates, without understanding the failure modes.

I'm just concerned about that. It seems we've got the cart before the horse, and are you going to get to this some time today, and if not, I'll raise it again toward the end.

MR. KURITZKY: Well, no, I can talk to it right now.

We are going to talk a little bit about it, particularly when I hit the WGRisk Activity, at the end.

But basically, I wouldn't say that we didn't come up with anything when we looked at failure modes.

As I mentioned before, BNL did a very detailed FMEA of the hardware for the digital feedwater control system, and so, there was a lot on failure modes there.

We recognize that it's not necessarily 100 percent complete. We don't know whether there are other failure modes out there that we haven't thought about.

But basically, unlike if you're looking for digital system assurance or software assurance, you want to make sure the system is adequate for licensing.

When we're looking to put something into a PRA, it's more or less trying to capture the reliability of the system, and the failure modes aren't the traditional failure modes that you might look at, if you're trying to totally analyze the system to make sure you're figured out every way it can go wrong.

It's more like, how does it fit into your model? Like, you know, we could have something as simple as software fails to generate the signal to initiate the safety injection system, and that may be the only failure mode that we have.

There isn't necessarily a whole set of failure modes that we have to go on. If you want to tell your -- if you want to increase your confidence that this system will do what it needs to do, if you're, let's say, in NRR and trying to improve a submittal on an application for a system upgrade, then you need to be more confident that all of the possible ways that system can fail, you've considered.

In the PRA, we're more interested in just how does it fit the model, particularly dependencies. That's one thing that -- that does kind of overlap there, is making sure that failures that occur in the system, how do they impact other parts of the plant, and that's obviously a very big concern with PRA.

But I mean, it's a little bit different arena there, and that's why when we talk about -- when we're going to talk about the framework that we're coming up for, for -- or taxonomy for failure modes, it's for PRA, as opposed to some of the other taxonomies that exist for other purposes.

MEMBER BLEY: The thing bothers me, and I see that things are going to be tested and we'll decide how many tests we need, by the PRA goals we have.

But if you don't know the failures modes, right, then you don't know if you're getting hits on failure modes that could be more troublesome than the ones you're getting hits on when you do the test.

Are there some hidden failure modes with dependencies that could mean now that we're having more integrated systems, that suddenly, here is that common cause failure we're worried about, and something that would take out all of our protection at the same time, because of some unusual conditions that exist, wherever.

And if you don't know the failure modes, how do you know that you're going to find what you're looking for when you do the test?

You're going to find the things that happen more often, sure, but are those the things we're most concerned about?

So, that is why I'm thinking if we don't really get a handle on the failure modes, we aren't going to pick up the key dependencies.

Now, you might -- you're talking about

dependencies across systems, and I'm talking about dependencies internal to multiple software systems.

CHAIRMAN STETKAR: To follow up on that, because, you know, you've tried to bring us back to the context of a PRA model, which is good because in practice, we need that context.

To follow up on Dennis' comments, though, there -- the current report that we're reading is focusing on a particular failure mode for a particular function.

This thing does not trip the reactor when the reactor is supposed to trip. Okay, that's one.

What about digital systems that monitor, for example, loading on diesels, pressure in the primary system, and determine whether or not I should start high pressure injection or low pressure injection, because I can only start one or the other, because I don't have enough capacity on my diesel, or shut off high pressure injection at a certain time, under a certain set of conditions, because the software then knows that the conditions are right for low pressure injection? There are systems out there that do that.

That is a failure mode that is analogous to what we call operator errors of commission in human

reliability analysis. The system does something that it's not supposed to do, at a time when you don't expect it to do it, because the conditions are just right for the system to make the wrong decision.

That is also a certain type of failure mode. If we're precluding our ability to address those failure modes in the research work, or precluding our ability to develop -- derive data for those particular types of failure modes, because we haven't thought about them, then it raises questions, as Dennis raised, about are we putting too much emphasis on collecting data, as we did 30 years ago, collecting all sorts of detailed databases for valve failure modes, before we really knew about what failure modes might be in a PRA model?

You know, packing leaks on one-inch valves versus packing leaks on two-inch global valves versus, you know, that -- people spend a lot of research -- resources doing that, until they discovered that they were essentially wasted resources.

MR. KURITZKY: I think --

CHAIRMAN STETKAR: I think you're going to hear this theme from us.

MR. KURITZKY: And it's a valid thing. I don't disagree, at all, but what I would like to put

out that is there we're not necessarily ruling out all that stuff.

What we believe is that we don't need to solve all that problem, before we can start moving forward on how we're going to model things.

So, in other words, we're going to -- like the example you gave, we can come up with a method to do a soft -- try and figure out how to calculate the failure probability of software to do the trip function, as you mentioned, but it doesn't mean that we -- we're also including these other functions that we have to identify and have to come up with ways to model those, also.

That's a piece of work that is ongoing. We're not doing this instead of that. You know, we want to do this things more in parallel, otherwise, the time line goes out very far.

So, once we know -- we know enough failure modes that we can say, "Okay, how can we model some of these failure modes that we know about," then as we discover more failure modes, then we have to see, do these methods still work for those, or do we have to come up with different methods to model those failure modes?

But I don't want to leave everything in a

serious type of arrangement, where we have a time line that just becomes unworkable.

MEMBER BLEY: In fact, let me be honest, I think my concern, and this has been over the last few years, is -- and I can understand, we want to do things in parallel, and my concern is, maybe we're not, and we're leaving the piece that could affect everything for the end, instead of working it hard all along the way. That is my concern.

MR. KURITZKY: Okay, well, all I want to say is, the work that we have would be -- the research with BNL is to pursue the quantification of a software failure, okay.

So, it's not that if we weren't having that, we would have this BNL team going through, trying to expand the failure modes, or the identification of failure modes.

That is work that occurs in Russ's group, and they're doing work both in-house and externally, to try and address that issue, and so, it's not a question of, well, since we have BNL doing this right now, they're not doing the other stuff.

I think there is -- there are different organizations --

MEMBER BLEY: I guess I'd just put in,

we'd like to hear from Russ sometime, not too far off, about what's going on, because --

MR. KURITZKY: And in fact --

MEMBER BLEY: -- we were going to ask him for that, for a long time.

MR. KURITZKY: And this very next slide I was going to go to, where I mention that there are number of other areas that still need to be addressed.

You know, right now, our focus is on software probability, but our previous research identified a number of other things that still need to be addressed, and of course, if we're going to have a robust modeling of digital systems in the PRA.

Some of these things are already being worked on by other groups, both in the NRC, for instance, Russ's branch. Other ones are being worked at by industry or by international organizations, and particularly, the stuff that's involved -- and Russ, you're going to hear about some of these things, in just, I think, a couple of weeks.

I think Russ's group is coming to talk to this Subcommittee in like two weeks, and so, they're going to talk to you about, I think work that may be being done in-house on failure mode identification or database development.

So, you're going to hear about those, also.

MEMBER BLEY: Yes, okay.

MR. KURITZKY: Very shortly.

CHAIRMAN STETKAR: Okay, I wasn't aware that that June 22nd meeting was going to delve into sort of those issues, but I think it's June 22nd.

MEMBER BLEY: Yes, it is.

CHAIRMAN STETKAR: Yes, right.

MR. KURITZKY: It's going to talk -- there is a number of -- and I don't want to put words in Russ's mouth, but it's going to talk about a number of related topics, and so, I think --

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: -- you'll probably hear something about those things.

CHAIRMAN STETKAR: Okay.

MEMBER BLEY: Yes, I thought it was a research plan, but maybe it is the research plan.

MR. KURITZKY: But that stuff is -- that's under Section 3.1.5 of the research plan.

MEMBER BLEY: Okay.

MR. KURITZKY: It deals with that.

CHAIRMAN STETKAR: Telegraphing for that meeting, if you can highlight, at least for our

Subcommittee, any work that's being done in particular, in the failure modes identification process, we obviously would be really interested --

DR. GUARRO: Can I add a comment --

CHAIRMAN STETKAR: -- interested in hearing about that.

DR. GUARRO: -- on this question of failure modes, because I think, you know, the failure mode question is very intimately connected to the question of what are you trying to identify and model, because you know, failure mode is one of those words, or actually phrase, that can mean a lot of different things.

Failure modes could be at the atomic level or at the system level.

So, I think what is important is that there is -- there is a criterion. If one wants to identify and look at failure modes, there should be some guiding criterion.

I would, you know, just put out there, a suggestion, which I tried to make other times, but I don't know if I've been, you know, successful in putting it in clear terms.

You know, when we talk about the inference that software can have on a complex system like a

nuclear power plant, we have to ask ourselves, you know, in practical terms, what does the software do?

The software -- you know, the word failure is the -- you know, the failures remain inside the software, and nobody would care. It is the interaction with the physical systems that makes it important, and how does software interact with the system, you know, in two ways.

Through the sensors, because that's where it gets the information that it processes, and through the actuators, which is -- that's how it makes things happen.

So, if we are to start anywhere, looking at failure modes, it should start from the interface failure modes, and then if you want to go down into the software, you can go down into the software, but at the very -- at the very minimum, I think that those two types of interactions determine what happens to the plant, because of what happens in the software.

CHAIRMAN STETKAR: Thanks. I think as we go forward, we need to be aware of the time here a little bit.

MR. KURITZKY: Have I used up my half-hour yet?

CHAIRMAN STETKAR: Not quite.

MEMBER SIEBER: You've got 20 minutes left.

CHAIRMAN STETKAR: The good news is June 7th ends at midnight tonight.

MEMBER BLEY: You've got plenty of time.

CHAIRMAN STETKAR: You've got plenty of time. Seriously, as we get into the details of actually, what you're doing, I think you're hearing from us that we'd like to at least have confidence that this particular effort that we're going to hear about today gives us confidence that indeed, we're not scoping the problem down so far that it may preclude consideration of either, in this particular area, data collection methods, or data processing, however you want to characterize them, data processing methods, that may be needed to support a more broad range of failure modes.

In other words, that we're not actually somehow de facto falling into that type of serial process that you very much want to avoid --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- simply because we scoping the current problem down so far that by default, it becomes kind of a series process.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: And with that, you can finish your half-hour, and we'll try to keep quiet.

MR. KURITZKY: Okay, just to quickly wrap up, the one point I do want to make, even with all these various gaps that still exist in the state-of-the-art, I do believe we will come up with a method for modeling digital systems in PRA, including software. We certainly can do it.

The lingering questions are going to be, is it going to be practical to apply these methods on a commercial scale? You know, it is one thing to do it as a big research project, and another thing to have every licensee PRA engineer be able to apply it.

And the second thing is, will the -- the results -- will the uncertainty on the results that this approach provides, will that uncertainty be so great, as to render that relatively meaningless in a regulatory environment?

I mean, we have to have enough confidence in those results that we can use them for some type of regulatory application, and that's another issue, given a lot of the uncertainties that are involved in failure mode completeness and the data that we would come up with.

So, those are the lingering questions that

are always going to be -- we'll have to keep in mind.

Okay, then real quickly, I'm going to wrap up, spending just a couple of minutes on this activity that I mentioned before, the WGRisk Activity that's currently ongoing on digital I&C failure mode taxonomy.

This effort is being headed up by the VTT Technical Research Center at Finland. They have the lead. The NRC is a very significant supporter and participant in that activity.

Its focus is on coming up with a failure mode taxonomy, or taxonomies, to support PRA, and again, this goes back to a little bit of what I was saying before, in response to Dr. Bley's comment, that we recognize there are taxonomies for different purposes, and we're really focusing here on ones that we can use for PRA, looking at both hardware and software.

And the process that we're -- that is often with these international activities, we're going -- we're scheduling a series of workshops, roughly two per year. We just had the first one a few weeks ago in Bethesda.

It was a very successful meeting. We got off to a good start. However, with all of these

activities, the key is whether or not we get the committed people to keep working on the issue, once they go back to their home desks, because it is the tendency for everybody to go back to what's been waiting for them, and they ignore it until the next workshop, and obviously, there is a lot of work that has to be done here.

It is going to have to keep going on throughout the interim periods between the workshops and so, it remains to be seen. The next workshop is scheduled for October, I think the third week in October of fourth week in October, in Finland, and we'll have to see what kind of progress has been made between now and then.

There is also two workshops going to be scheduled for 2012. We have one, I think, in May, in Germany, and the other one hasn't been scheduled yet.

The overall schedule is to get the best practice guidelines put together and submitted to CS&I by probably mid 2013.

CHAIRMAN STETKAR: Alan, I'd promised we'd keep quiet, but it's impossible for me to do.

So, you mentioned that this effort is being headed by VTT?

MR. KURITZKY: Yes.

CHAIRMAN STETKAR: And that's under OECD?

MR. KURITZKY: Yes.

CHAIRMAN STETKAR: Okay, are -- is anyone from Britain involved in this, that you know of? Are the British --

MR. KURITZKY: Nobody from Britain -- oh, that's a good point, speaking of Britain.

We do have people from the Center for Software Liability and the City University of London.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: Bev Littlewood and Lorenzo Strigini, et cetera, we have them on subcontract with BNL, to be our expert consultants on the work that Louis and Meng are going to describe for you, in the next few presentations.

We've been bouncing stuff of them, to get their input on that.

But as far as this particular workshop, we did not have anybody from England. In fact, we had a previous workshop similar, or activity, on digital I&C modeling back in 2008.

It was essentially, culminated in a workshop in October 2008, in France, and again, there we really wanted to get the British involved, because they have a lot of experience with Sizewell, and

unfortunately, the people that I talked to over in Britain were not able to come.

They had nominated other people to come in their place. They did not show up, either.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: So, we really have not gotten --

CHAIRMAN STETKAR: The reason I asked is, I'm somewhat familiar with -- I have to be careful of how I characterize this, British statements of their experience and developments in the area of software modeling.

I haven't personally seen any of that work, and I was just curious because they tend to say they've thought a lot about the problem and worked on it, in particular, you know, in a realistic arena of Sizewell, whether or not they were actively participating in this effort, and you answered that question, so, thanks.

MR. KURITZKY: Right, yes, we've been trying to get them involved, but they just haven't --

CHAIRMAN STETKAR: Fine, fine.

MR. KURITZKY: -- gotten to the point. So, basically, that's -- I guess my half-hour is just about up, that's basically all I have.

CHAIRMAN STETKAR: Just about. Any other questions for Alan, because I think what I'm going to do is, before Louis gets spun up and started, I'm going to move the break up, and take a break now.

So, if any of the members or anyone has questions for Alan, let's get those out.

DR. HECHT: I have a question and that is, is the goal of this work -- I'm having tremendous problems with failure modes, because I think I know what they are, and I think there are only five software failure modes, and what we're really talking about is effects of various kinds.

So, my question to you is, is the objective of these failure modes to come up with base events that fit into fault trees, putting it into very concrete terms?

MR. KURITZKY: Are you talking about the WGRisk Activity of the research that we're --

DR. HECHT: No, I know the research today, that you're doing.

MR. KURITZKY: Okay.

DR. HECHT: I'm talking about the activity that you were mentioning just now, the VTT work.

MR. KURITZKY: Okay, that is to come up with failure modes, I think really, to plug into a

PRA, to plug into --

DR. HECHT: So, it's base events or primal events, whatever you call them, that go into a PRA?

MR. KURITZKY: Right.

DR. HECHT: Not necessarily the same thing as what I would call a failure mode.

MR. KURITZKY: Right, and that's why I said there is -- that's why there is different failure mode taxonomies that exist, and people have them for different reasons.

But just real quick on that, it's what failure modes to go into the PRA, but we're also looking at different levels, because we don't know what the modeling approach will necessarily be --

DR. HECHT: Sure.

MR. KURITZKY: -- and different PRAs may be at different levels, or -- so, we're trying to do it in a few different levels, but it's essentially all of them that would, at some level or another, go into a PRA.

DR. HECHT: Okay.

CHAIRMAN STETKAR: Good, with that, we will recess until 10:30 a.m.

(Whereupon, the above-entitled matter went off the record at 10:12 a.m. and resumed at 10:30

a.m.)

CHAIRMAN STETKAR: Okay, let's reconvene and we'll hear from Louis, I guess.

MR. CHU: Yes, I'm presenting the selection of quantitative software reliability methods.

Actually, in Alan's presentation, he already touched upon many of the things I have in the presentation, except my presentation, I give you a little more detail.

CHAIRMAN STETKAR: Can you move the microphone a little bit closer to yourself? Just be careful that you don't hit it with your paper and things.

MR. CHU: Okay.

CHAIRMAN STETKAR: To make sure we pick you up a little bit.

MR. CHU: To give you a little background of the earlier work, and I'll show you the list of the desirable characteristics that we developed, and show a summary of evaluation of the QSRMs and how we select the QSRMs, and Alan already gave you the bottom line answer.

These two earlier studies can provide some background of our current work.

First, the workshop of the experts. In that workshop, we have experts, international experts and professors and our NRC staff. We had industry representatives participating in the workshop, and during the workshop, we established a philosophical basis for modeling software failure, and I'm not going to read the whole description.

But basically, it says software failure is somewhat similar to tossing a coin. People say software failure is deterministic.

From the same argument, you can say tossing a coin is deterministic, if you were to -- able to account for everything that affects the movement of the coin.

On the other hand, because our inadequate knowledge about software failure, or tossing of a coin, we use a probabilistic model to represent -- to model software failure.

In that workshop, we also have experts discussing modeling methods and quantification methods.

In our -- the second study on the slide, the BNL report, we developed a set of desirable characteristics and we performed a review of these methods.

The only thing that we didn't do in that study was evaluating those methods against the desirable characteristics, which is done in the current study and presented.

Of course, that earlier review provided background information. We had provided comments on those methods, and these are information that provide the basis of our evaluation of the methods against these characteristics.

CHAIRMAN STETKAR: That second bullet, the BNL-94047-2010 report, that report is the basis for the conclusions that are documented in this draft NUREG, is that correct? That's the evaluation of the methods, with respect to the desirable characteristics?

MR. KURITZKY: That report talks about the methods and identified the characteristics.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: This report actually compares them now against the characteristics.

CHAIRMAN STETKAR: Which report?

MR. KURITZKY: There's a formal comparison on the other one. I mean, but we talked about them in relation to that, but not like comparing the different methods to each other and to the characteristics.

CHAIRMAN STETKAR: That's -- let me understand which report.

MR. KURITZKY: I think that's chapter two.

CHAIRMAN STETKAR: It's chapter two, chapter two, though, if I read it in the current report, is essentially a summary of the conclusions from the evaluation.

It doesn't strike me as a very thorough evaluation, quite honestly.

MR. KURITZKY: Right, well, because the bulk of the work was done in the previous report, where we looked at the various methods and what they could accomplish, and we identified the characteristics, but we didn't do the step where we said, "Okay, how well does this method meet those characteristics?"

CHAIRMAN STETKAR: I understand, but what I'm trying to get my hands on is, as I'm familiar with the former report, that identified the characteristics and the history of that.

I've read the draft NUREG, whatever chapter it is, two or three or --

MR. KURITZKY: Two.

CHAIRMAN STETKAR: -- that from my perspective, summarizes the conclusions from the

evaluation of each of the methods, with respect to those characteristics.

It seems to cite a separate report that actually describes that evaluation. Is there a separate report that describes that and the --

MR. KURITZKY: No, no, there isn't.

CHAIRMAN STETKAR: There isn't?

MR. KURITZKY: And chapter two is somewhat light. I mean, it doesn't have as much detail in it.

CHAIRMAN STETKAR: It is really light.

MR. KURITZKY: And that's one of the --

CHAIRMAN STETKAR: The reason I brought it up is, I think -- I was led to believe that there was more detailed supporting report, probably a Brookhaven technical report, that more thoroughly described the evaluation process, you know, why does this particular method not quite satisfy this particular characteristic? Therefore, we draw the following conclusions.

I guess what I'm hearing is there isn't that, because my question was going to be, I'd really like to see that evaluation, you know, and why isn't it included, for example, as an appendix to this draft NUREG, because that's very valuable information.

MR. KURITZKY: Right, and unfortunately,

it hasn't been documented yet. We got that comment from management, too, on the same report.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: So, in the following report, there is going to be more information --

CHAIRMAN STETKAR: Oh, good.

MR. KURITZKY: -- on the --

CHAIRMAN STETKAR: I would sure hope so.

MR. KURITZKY: Okay, yes.

CHAIRMAN STETKAR: Okay, thanks.

MR. KURITZKY: Point well taken.

MR. CHU: But a lot of the technical evaluation is already in this BNL report, the second bullet on this slide.

CHAIRMAN STETKAR: Okay.

MR. CHU: It's just making use of those technical information and review, to support the evaluation against the characteristics that we -- that's what we did in the report that I'm presenting today, that we are presenting today.

CHAIRMAN STETKAR: Okay, but I think, as I said, I think the briefing that we had last -- a year ago, summarized the efforts that were done to identify the characteristics, and it provided a reasonable discussion, if I recall correctly, of the

characteristics and some of the -- you know, the bases for why those are a reasonable set.

So, you know, I thought I had a fairly good understanding of the context of those characteristics.

What I seem to be missing is the details, you know, the actual documented thought process of the evaluation of each of the -- however many there were, half a dozen or so --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- identified methods, against those characteristics.

MR. KURITZKY: Right, and like Louis said, some of that information was in the other report, but this report needs to beef-up in that area, and it's going to get beefed-up in that area.

CHAIRMAN STETKAR: Yes, I think it really does.

MR. KURITZKY: Yes.

MEMBER BLEY: But that last Brookhaven report, last year Brookhaven report, do we have that one?

MR. KURITZKY: Yes, we should have them all. I'm not sure if last year's meeting -- last year's meeting was before that was a final report.

But I think --

MEMBER BLEY: Because Christina, would you make sure we have that?

MR. KURITZKY: Yes, we can forward it to you, but I --

MEMBER BLEY: Okay, I'll look around a little b it here, maybe I'll find it, I'm not positive.

MR. KURITZKY: Okay, the 94047.

MR. CHU: Okay, about development of these characteristics, it was mainly done in the BNL report, that Dennis was asking about, and they were developed based on our team's experience in doing literature research and based on our understanding of what PRA needs.

The purpose of the characteristics are to identify promising methods for QSRMs to support PRA modeling of digital systems.

This slide lists the 10 characteristics. I just want to find out a little bit, in the previous version, in that BNL -- earlier BNL report, we have nine characteristics.

In our current study, we added the tenth one, on whether or not data needed, in supporting the analysis, exists and can be collected, and I'm not

sure I need to go through this desirable characteristics one by one. Maybe I'll try.

First one, the description of the method and application is comprehensive and understandable. I think this is typically set by many of the methods.

The second characteristic, the assumptions of the method have reasonable basis. This is a big question, of course, how you assign an answer -- assign the assessment objective, because it's a broad-based characteristic.

The third one, the method allows for consideration of specific operating conditions of the software. We recognize that software failure is sensitive to context.

An earlier ACRS meeting, we pointed that out, before we include this desirable characteristics.

The fourth, the method takes into consideration quality of life cycle activities. We know that the software development activities affect the quality of software. Therefore, it's desirable that the quantification method can prove this, and as Alan has pointed out, the BBN methods supposedly is the only method that -- looking to the quality of life cycle activities.

The fifth one, the method makes use of

available test results and operating experience. The sixth one, the method addresses uncertainty.

The seventh characteristic, the method has been verified and validated. The eighth one, the method is capable of demonstrating the high reliability of safety-critical system.

The ninth one, the method should be able to estimate parameters that can be used to account for software common cause failure.

The tenth one, I had already mentioned, but you can see that the characteristics cover many issues and it's not possible that any existing method would satisfy all of these.

So, I guess that's why they are desirable, and we try to use these characteristics to help us to select promising methods.

DR. HECHT: Louis, if I could ask a question about number four, which is the life cycle methods, or the quality of the life cycle.

I have two comments I wanted to make on that.

Number one is, do you think -- do you envision that software going into a nuclear power plant, either as the safety or the control function, would be developed with anything less than very high

quality software development methods?

MR. CHU: I would expect they will -- I would expect safety-critical system to satisfy the NRC deterministic requirement.

In that sense, they should meet the high quality requirement.

DR. HECHT: Right, so, to that extent, it becomes kind of the same across -- it's a constant, so, it's a constant, kind of like using, I don't know, binary code for representation of software commands.

In other words, what I'm saying is, does it really make a difference, because it's constant throughout all of the systems that you would be using -- you would be analyzing, or under consideration.

MR. KURITZKY: Let me just jump in for a second. I think it doesn't make a difference if you're looking at one application versus another one that's going into the nuclear plant for a safety related system.

But it is important to -- when you come up with, let's say, the base line failure probability, or distribution for failure probability of the software, that the experts who are going to have to help us rate those probabilities recognize that they're going to have that high level -- that high quality of the life

cycle activity.

So, you're right, it should be a constant amongst the software, but it's going to be a part of that ultimate, in our case, our theory distribution of what that software failure probability is.

So, yes, I would expect that that's one where if we can get -- and I think when Louis goes to talk about this a little more, there is a generic aspect to that, that once, you know, we set up this generic BBN, based on the -- for instance, what's in Branch Technical Position 14 on the life cycle for all the activities for the software, that we should get that same essential value, or contribution from all of the different software that would go into that nuclear plant.

DR. HECHT: Well, that was going to relate to my second question, which is, given that we have a base line that pretty much defines that there is going to be trace-ability and documentation and testing, with respect to what's been traced, that's basically -- those are basically common characteristics of all the methods.

How would anybody assess what effect that has on failure probability, even on an expert level?

I mean, I know that high quality

development methodology is necessary, but not sufficient.

MR. CHU: I think when we develop our model, we want to have the capability of differentiating different qualities.

In terms of your example, say traceability or completeness or documentation, you know, we tried to come up with some kind of questions on how well these activities are implemented, so that the questions can retract different degrees of quality.

In that sense, in one application, you know, you may not do as good of a job, than another application, and the model can somehow differentiate them.

CHAIRMAN STETKAR: Well, I want to see if we can move along here, a little bit, and the only reason I say that, I don't want to necessarily shut off discussion, but in some sense, this is a bit of history because in our last year's presentation, with the exception of number 10, I think that we went through a lot of these characteristics --

DR. HECHT: Okay, well, I --

CHAIRMAN STETKAR: -- and we had some discussion of them, at that time.

DR. HECHT: I can raise it again, part of

the BBN thing, but --

CHAIRMAN STETKAR: That's fine, how effectively, you know, the proposed quantification methods here, you know, may address some of these characteristics.

But I don't think it's appropriate for this meeting, to enter into the discussion about the relevance of this set of 10. That's a bit of history.

DR. HECHT: Just, and when you hit slide 14 of the BBN approach, it's --

CHAIRMAN STETKAR: Well, we'll deal -- I think, bring it up again in the context of, you know, a particular proposed method.

DR. GUARRO: John, just with exception on number 10 --

CHAIRMAN STETKAR: Number 10 is something they added in the last year.

DR. GUARRO: Okay, because I had a comment on number 10, real quick.

CHAIRMAN STETKAR: Okay, real quick.

DR. GUARRO: Which is perhaps, you know, that 'and' should be an 'or', because the data and new information exists and can be collected, you know.

To pose, you know, the requirement that the data exists in ready-made format, when you're

really in the development stage and nobody has done this type of work before, could be used, you know, to exclude stuff that instead, should be done, okay.

Now, that's my comment. Nobody has formulated the software failure databases in a ready-made useable format. So, you know, it is the job of the method, to define the form that is needed. That is my comment.

MR. KURITZKY: Just to clarify, I think the point there is not that it necessarily can be collected, like right now, we can go out and we can go get it.

It's more like, it potentially can be collected, something that we could theoretically get.

DR. GUARRO: And that's what I -- because that is what I hope you meant, but I just wanted to make that point, clear.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: Point well taken.

MR. KURITZKY: Well, theoretically.

MR. CHU: Based on our review of the methods, we found these common limitations of the methods.

The first bullet, test profile versus operational profile. Many of the methods use data.

It could be operational data or test data. In particular, when you use test data, the issue applies.

How realistic is the test profile? How well it represents the real operating condition of the system?

Second bullet, context specificity. Again, it's the issue of context and most methods do not specifically consider the context.

I guess, an exception would be Sergio's work in the CSRM, where you use a PRA model to identify the context for testing.

In our later presentation, you know, we talk about statistical testing, we're using the same idea, basically, use PRA to define the context.

Demonstration of high reliability. That's another issue that's hard to demonstrate, due to different uncertainties. For example, if you do a test, how realistic is your test profile? By how much -- how much error can you introduce?

Failure mode specific modeling, the methods we look at do not consider different failure modes. They just have system failure.

Common cause failure of two diverse digital systems, I guess this issue relates to the earlier question about considering new reactors, and

we heard that for EPR design, they have -- they don't have identical channels.

Instead, the channels may be different, have different platform or different software, and in that sense, it makes it more challenging for considering common cause failures.

If the -- I think people tend to agree, if you have redundant identical channels running the same software from PRA to the point, we assume, you know, the channel will fail together.

So, common cause failure treatment is simple. But now, if you have channels that run different software or run on different platforms, then how do you account for the common cause failure?

But in general, the QSRMs do not address common cause failure. An exception may be at the end version programming, which is -- that we -- not yet application on common cause failure-related issue.

DR. HECHT: Excuse me, Louis, I'm not sure I understand that last point.

You use the word system there. Isn't it too much to ask, if you're only looking at a prediction method based on the software, that -- or I should say the software perhaps executing in isolation, that you consider the system?

MR. KURITZKY: Well, first of all, let me say that probably there should be more accurate digital sub-systems or systems.

But it raises -- it raised the issue or -- or addressed the issue that Louis mentioned about, for instance, the one design where they were going to have -- they were contemplating having separate software, separate platforms for different channels within the system.

And so, then you still have to be able to address common cause failure of supposedly, two diverse systems -- sub-systems or channels.

DR. HECHT: Meeting the same specification?

MR. KURITZKY: I don't know, I assume that they would meet the same -- I would assume that it's part of the same system, it would have to meet the same specifications.

DR. HECHT: Okay, so, that's one issue because then certainly, of course, the common cause failures could be caused by the specification.

MR. KURITZKY: Right.

DR. HECHT: And so, once again, not fair to impose that on the software.

MR. KURITZKY: Well, it's not that we're

trying to impose -- that it's unfair to impose on it. It's just an issue that has to be addressed.

In other words, when we're going to model that in the PRA, ultimately, we're doing this to include it in the PRA, and if such a design exists and is to be incorporated into the PRA, then we have to be able to address the issue of -- you know, we don't -- as Louis said, we don't want to assume a failure -- a common cause failure probability of one, which we might have to, if it was -- which we might do for simplicity sake, if it was the same software on both channels, or both sub-systems.

DR. HECHT: Right.

MR. KURITZKY: But in this case, we wouldn't want to assume it was one. That would be overly conservative, but on the other hand, we don't want to assume it's zero either, because there are aspects which exhibit the same specification that would lead you to have some probability that the two could fail.

DR. HECHT: So, then in that case, that point is really the common cause failures of two diverse implementations, because maybe that's a better way of saying it.

MR. KURITZKY: That is probably better

terminology, yes.

DR. HECHT: Right.

MEMBER BLEY: We're going to be throwing these words around all day, and Alan started by saying there are some people that say software doesn't fail.

I don't think I found in the document anywhere, a definition of what you guys mean by software failure and software failure rate.

MR. KURITZKY: You know, not in that document. I think that the BNL report on the philosophical basis, I think there is one where they have a definition of software.

MEMBER BLEY: We've tossed a lot of things around, as I recall.

MR. KURITZKY: What is that?

MEMBER BLEY: As I recall, that tossed around a lot of concepts.

MR. KURITZKY: Yes, there were a number of concepts, but I think software failure -- we define it -- software failure --

MEMBER BLEY: I just think that it would be appropriate, I think, to say in here, exactly what --

MR. KURITZKY: Echo it in the report, yes.

MEMBER BLEY: And after lunch, maybe we

can come back to it, and tell me exactly what we mean.

DR. HECHT: Is it divergent from the specification of the system, due to a software issue?

MEMBER BLEY: I'd like to hear what they have to say.

MR. KURITZKY: Yes, well, the thing is, that's going to be something where everybody -- you can poll 20 software reliability guys or whatever, and you're going to get 20 different definitions of software failure.

CHAIRMAN STETKAR: That's fine, you know, we're dealing in an area where, you know, if you ask 20 different experts, you're going to get 20 different very, very strong opinions, obviously.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: At some point, you know, in the research program and practice, there needs to be convergence and you know, one way to achieve that convergence is through leadership.

So, that is why I think why Dennis asked, you know, regardless of what the other 19 experts might --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- consider is -- because we'd like to clearly, see your definition.

MR. KURITZKY: Right, we'll put our hats --

CHAIRMAN STETKAR: Because you're working, you know, this particular research is focused on that definition.

MEMBER BLEY: And I don't know how someone carries out the research you're proposing, unless they really know what they're counting.

MR. KURITZKY: Right, well, I think we have a pretty strong understanding of what we mean. It's more semantics than terminology, quite honestly.

CHAIRMAN STETKAR: And that's right.

MR. KURITZKY: Right.

DR. GUARRO: Yes, in fact, I think in that respect, at least it's important to just put out there, what you mean, by certain things, whether everybody else agrees or not --

MR. KURITZKY: Right.

DR. GUARRO: -- you know, you have no control over that.

MR. KURITZKY: Right, for the final report, we'll put it in. Thank you.

CHAIRMAN STETKAR: And if we could, I mean, honestly, if -- during the lunch hour, if you could -- you know, develop some information, not in a

slide form --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- but at least orally, kind of elaborate on what that particular context is, we'd appreciate that, since it's available someplace.

MR. CHU: This slide shows SRGM that we evaluated, reliability growth method is the most popular method, and you see in your literature, BBN method, this one -- some few applications in the nuclear area that you will hear more about, later in the presentation.

MEMBER BLEY: You said the most popular of the growth methods. Aren't they primarily used by developers, to improve the development process? That's why they're popular.

MR. CHU: Yes.

CHAIRMAN STETKAR: And aren't they similar to sort of the environmental qualifications process that people use? They give you some measure of confidence that the, in this case, software or the environmental qualifications, that the -- the item, if I can call it that, meets some minimal level of reliability that's used as sort of a warrantee for the equipment, or qualification in a licensing perspective.

They don't really give you much information about what the actual failure rate might be, or certainly, any uncertainty about that failure rate. Is that a reasonable characterization?

MR. CHU: They do estimate failure rate. That's for sure. But how realistic is that estimate for, you know, the software that's used in its actual application is kind of questionable.

CHAIRMAN STETKAR: Okay.

MR. CHU: It's probably -- one reason is, people seem to develop many variations of the method, different models, under this category and people write a lot of papers about it.

But in actual application, we actually have not seen any recent real application of the method. We've seen papers, recent papers that use data from -- collected from 20 years ago, that seem to say, you know, there is a new mathematical model, using data from 20 years ago, but what happened during those 20 years, are there more data, more applicable data.

CHAIRMAN STETKAR: Yes, it's easy to write papers about methods. It's not so easy to collect data.

DR. HECHT: Your point is well taken. The

key characteristic of the software reliability growth models that's -- there is a defect removal process ongoing --

CHAIRMAN STETKAR: And at some point --

DR. HECHT: -- and an operational, which implies that the software is undergoing change. You don't change an operational system and therefore, what you'll get is, you'll get the terminal reliability at the end of the development and you --

CHAIRMAN STETKAR: Yes, at which point -- this is good enough for the thing that we're going to sell to you and you, as a customer, say, that's good enough for me.

DR. HECHT: Yes, but it is -- you're correct, that it's not really appropriate to use in an operating system.

CHAIRMAN STETKAR: All right.

MR. CHU: The third bullet, test-based methods, again, we'll have more presentation later.

Frestimate is a method that uses regression analysis and data collected from past software developments, to estimate the defect density, or the number of defects in the software, and then the method converse that estimates number of fault into a failure rate.

The difficulty with this method is that the documentation of the method is available, it's brief, and we don't know the data that's -- there is no information about the data past software development activities that we heard about.

Some project may be development of software for ATM machines, but in general, attempt to change the method, you know, use regression analysis is a reasonable way of doing that, but where is the data? That is the question.

CHAIRMAN STETKAR: Louis, in terms of that, just because I'm not familiar with the method, when you say use of regression analysis to estimate the number of defects in the software and then produce information about an estimated failure rate, you mentioned an ATM machine, and I was going to use the analogy of an automobile, but I'll keep the ATM machine.

Have you seen anything that looks at -- if I can use the term failure rate, for particular failure modes.

So, for example, if I can envision my walking up to an ATM machine, one failure mode that I might be interested in is, the ATM machine eats my card, and never gives it back.

DR. GUARRO: I have one little point on that.

CHAIRMAN STETKAR: Okay, all right, as we all do. Another failure mode is, it takes my card, debits my account, but never gives me any money, and another failure mode is, I put my card in and it starts spewing out $100 bills, faster than I can particular --

MR. CHU: Okay.

CHAIRMAN STETKAR: Those are three different failure modes, all of which might be attributed to certain defects in the software, but is this just a generic -- any one of those failure modes might occur, and I need to worry about them, or is it focused -- okay, you don't have that information, okay.

MR. CHU: It's just from the failure rate.

CHAIRMAN STETKAR: Okay, fine, fine. You know, I've read what you said about it in the report. I was just curious, whether it was any more focused than that.

MR. CHU: The information is proprietary.

CHAIRMAN STETKAR: Okay.

MR. CHU: I guess if you ask them to apply the method, they can do it for you, but you don't know

what's in their black box.

CHAIRMAN STETKAR: Okay, thanks.

DR. GUARRO: John, just for information, I have one little point also, on the application of Frestimate to an analysis, in which essentially, done that way versus using historical data of mass emissions, in cases in which there were software failures.

There was essentially two orders of magnitude difference.

CHAIRMAN STETKAR: Okay.

DR. GUARRO: But the first one being the more optimistic --

CHAIRMAN STETKAR: More optimistic, okay. You know, we've heard, in the previous -- in the meeting a year ago, kind of similar conclusions, and I was just curious whether we'd learned anymore about that, but it doesn't seem to be, for whatever reason, whether or not the methods are not reasonably available or whether they're not really responsive to the types of applications that we're interested in.

It doesn't seem to be a pragmatic methodology. I was just curious, you know, how much more you've examined it within the work that you've done over the last year. Thanks.

MR. CHU: The metrics-based method is the work done by Carol Smith. We look at earlier study that she and her students performed.

We look at two methods. We look at two of the -- several methods in her report. One is the mean-time to failure method. The other is different density method, and a way of looking at it is mean-time to failure method effectively is saying you collect data and estimate the failure rate of failure probability.

So, that is a standard method of different density method, we can look at it as a white box method, in the sense that you actually -- they develop a model for the software and go through testing or go through inspection, to identify faults and use the identified faults in their model, to estimate software failure probabilities.

Standard based method, basically, we're talking about the IEC-61508. There, they specified deterministic requirements or recommendations on how you develop the digital system, and they assign some kind of failure rate or failure probability, for different cells, and the issue is, there is no strong basis for those assignment of those numbers.

Of course, this difficulty is common to

many matters, when you hear about the Bayesian belief network analysis. We have similar type of difficulty.

Yes, other methods, the CSRM, the work done by Sergio that I mentioned, N-version programming is some method that specifically look at having different teams of software developer follow the same specification, develop software separately, and the conclusion of the experiment is that they do fail, different version do fail together in some cases, but in other cases, they don't fail together.

So, in a sense that, in PRA terminology, there is a beta factor you can use to represent the degree of dependency between different versions of software.

So, you don't completely eliminate the dependency, but this is only one aspect of diversity. I know, Oak Ridge had the report on different diversity strategies, implementing digital systems, and in general, common cause failure modeling issue will try to look into different diversity -- all the diversity strategy implemented in a digital system.

This slide summarizes the evaluation. It was done mainly, by me with Alan's review. We assigned different colors. The dark color means 'yes'. The grey color is 'maybe', and white color is

'no'.

So, the more dark colored answers, the better is the method, and as I commented earlier, we probably want to provide some more documentation on how we arrive at this evaluation.

MR. KURITZKY: I think, let's go -- go back to that bigger, just to point out.

One thing that we can see from that figure is, there is no particular single method that jumps out, head and shoulders above everything else. We didn't stumble on the Holy Grail and say, "This is the definite method to pursue." If you could just stay on that figure, go back to that figure.

And the point is, so, when we selected a couple of methods to move forward on the BBN method or the statistical test based method, it's not that we're saying the other methods are not adequate, and so, and these are the only two that are worth pursuing.

There are a number of methods out there. They have pluses and minuses, and for specific reasons, we picked the two that we're going to pursue, but it's not to say that the other ones don't have possible merit, also, and some of them are actually being pursued by other people.

Even within the NRC, there is work, for

instance, on the metrics method that's being done under Russ's organization.

So, I mean, we don't want to give you the impression that we've picked the BBN and the test based method, black box testing, and that the other methods, which is white box, I think, don't have any value.

They do, and they could be pursued, it's just that for various reasons, we've particularly pursued the two that we've chosen.

CHAIRMAN STETKAR: But again, you got the point. I think it's important, this focuses, you know, your efforts, going forward, that the thought process, that if it was, you know, you and Louis, that's fine, you know.

I think it's valuable to capture and document that thought process.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: You know, we made the conclusions, this is grey, rather than black, because.

MR. KURITZKY: Right, and then, in fact, that's something we -- we've already had that noted.

CHAIRMAN STETKAR: Yes.

MR. KURITZKY: Like I said, our management review identified the same thing, and we already have

a note for the final report.

It was just too much -- unfortunately, to get it out in time, would be difficult and --

CHAIRMAN STETKAR: The problem is, that's an important part of that ranking process.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: And if it isn't documented, it's lost.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: You know, if we have this same meeting a year from now, and perish the thought that you and Louis disappear for some reason -- yes, that's right, and he will say, "Ghee, I wasn't really involved in that." We need to go back -- we need to do that whole effort --

MR. KURITZKY: Right, yes, so, I would -- it will be in the final version.

CHAIRMAN STETKAR: Okay.

DR. HECHT: Can I ask a question about, on the previous slide, on nine, which is basically the common cause failure probabilities, you say that none do, and I'm just wondering why you don't feel that test based methods work.

In other words, if you perform tests and you see that there is failures on -- of one

implementation that don't affect the other implementation, then wouldn't you be able to make some measure of -- some assertion about the correlated failure probability?

MR. KURITZKY: Well, I think if we were doing testing right now, at least as this -- through the merit phase, I think we would be testing one implementation.

So, that's the point. That's like another advanced level, would then be to try and test a combined -- you know, the combined implementation, or you already have two diverse implementations, and testing them both.

DR. HECHT: Right.

MR. KURITZKY: Right now, we would just be testing one implementation.

CHAIRMAN STETKAR: And I think you have to be a bit careful, if I heard you right, Myron, that confidence in how big something is not, let that sink in, does not necessarily give us about -- uncertainty about how small it might be.

So, just simply the non-observation of things that might be coupled, gives you some confidence about how infrequently those types of failures might occur, doesn't really give you much

estimate -- much information about how frequently they do occur, or what the causes might be.

MR. KURITZKY: And given that they would expect to be hopefully, very low likelihood --

CHAIRMAN STETKAR: Well, but again, in risk assessment, you know, we need an uncertainty distribution and a mean value estimate, for that.

MR. KURITZKY: Right, but I mean, going back to what you were saying, the fact that we know it's not above this, doesn't really help us know how low it is, and in this case, it's something that would theoretically be very low, so, you know, it's --

CHAIRMAN STETKAR: You need whatever it is, yes.

MR. KURITZKY: Right.

DR. HECHT: Well, you could make the assumption -- some estimate of the upper limit, with the degree of confidence, couldn't you?

CHAIRMAN STETKAR: Upper limits don't help us much in risk assessment.

DR. HECHT: Okay.

CHAIRMAN STETKAR: They some --

DR. HECHT: Sometimes.

CHAIRMAN STETKAR: I'll just leave it at that.

DR. HECHT: All right.

MR. CHU: Alan kind of covered some of these slides.

MR. KURITZKY: I'm sorry.

MR. CHU: As I mentioned, you know, now, it matters that satisfy all the desirable characteristics, and in our evaluation, we just identified two methods that we considered more promising.

It doesn't necessarily rule out the possibility of using other methods.

Our selected methods, are kind of consistent with the opinion of experts at the expert workshop we had, and are consistent with the National Research Council, also.

Of the two methods, we'll have more detailed presentation. BBN methods, the advantage of that is its ability to capture the quality of software development activities.

On the other hand, there is difficulty in converting qualitative information into quantitative information, in particular in software failure rate or failure probabilities.

Test based method provide you kind of measurable results that you can use to try to quantify

software failure probability. Of course, it's still subject to different simplifying assumptions involved in the testing that -- in the later presentation, we'll have more discussion on testing and BBN method.

This is all I have to say on the selection models.

MR. KURITZKY: Ready for the next presentation?

CHAIRMAN STETKAR: Yes.

MR. KURITZKY: Okay, so, why don't we switch over here?

CHAIRMAN STETKAR: Meng, we need to break at noon for a couple of other reasons. So, I don't know if you can get through this in 40 minutes. I don't want to rush you, but keep aware of that fact, and if you have a convenient break point, somewhere in the 30 minute to 40 minute time frame, keep that in mind, so that we can break at a convenient point, because we do have to stop around noon.

MR. YUE: Okay, I'll try.

CHAIRMAN STETKAR: And you know, I don't want to rush you, to try to get everything done in 40 minutes, but just keep that in mind.

MR. YUE: Okay.

CHAIRMAN STETKAR: If you could, thank

you.

MR. YUE: Okay, so, BBN method, or Bayesian belief network model is one of the selective methods, as discussed by Louis just now.

We are going to present some details about this. The first part is always the one about some basic knowledge about BBN network.

Then we are going to discuss environment of BBN model, how to use it to estimate a software failure probability.

Like every method, BBN method also has these issues and limitations, some of them were already mentioned by Louis and Alan, just now.

Uncertainty is intrinsic to any method. It needs to be probably addressed. We will present a little bit of thought about how to treat the uncertainty.

We've developed a procedure for coming up with BBN model, for example system, and will be presenting that and also, one more slide will be -- present the status of the BBN modeling.

Some basic stuff about BBN network. It's a probabilistic graphical mode. It consists of nodes, and these nodes, they are connected by their active edges.

Each may represent a random variable and their influences are reflected by the directed edge, or even you can -- we can call them conditional independents.

A basic assumptions encoded in BBN network -- in BBN is giving its parent, that is this node is conditionally independent of its non-descending nodes.

In other words, all you have to know is about the parent node. Then you can infer information about this node. You don't need to know other non-descendent nodes.

Because of this conditional independence, the joint distribution of BBN model can be formulated in this -- this is very complex -- it's a relatively more compact form, and the Bayesian inference can be performed by updating this joint distribution, using acquired evidence. There are a set of software tools that can perform this.

This is a very simple BBN example. It has four nodes. All of them are binary. They have two states, true or false, and it might be cloudy. The cloudy weather may affect the chance of turning on sprinkler, also, being cloudy may affect a chance of raining.

The fact that the grass is wet may be due

to two causes. One is, sprinkler is on. Another one is, it is raining, and from this simple network, we can perform inference or queries, and the two types of inference, one is called bottom-up inference.

Basically, you are diagnosis, the causes, for example, given the fact that the grass is wet, we want to find whether it's more likely due to the sprinkler is on or because of it is raining. From this model, we can perform this kind of inference.

Another inference is called top-down inference, or forward inference. Basically, it's going to predict the effect. Given the fact that it is cloudy, how it's going to affect the probability of the grass to be wet.

The strength of the BBN network is probability of capturing the subjective opinion from experts and also, combining the subjective opinion with disparate information, both it can be quantitative and qualitative.

Eventually, qualitative information, they have to be converted into quantitative information, for the BBN to -- for quantifying the BBN network.

CHAIRMAN STETKAR: Meng?

MR. YUE: Yes.

CHAIRMAN STETKAR: Before you leave that

slide, I just want to make sure that I understand.

You kind of characterized this as the BBN can be used either from a bottom-up to look at the most likely contributing causes to an observed condition --

MR. YUE: Right.

CHAIRMAN STETKAR: -- or from a top-down, and sort of predictive mode.

I'm assuming that for the particular application of software failures, we're invoking that top-down sort of thought process.

MR. YUE: Right.

CHAIRMAN STETKAR: Kind of in accordance with our theme, it's fundamentally important to understand that what I'm interested in, ultimately is wet grass and not, you know, my child ran outside, for example, so, that before I can construct one of these BBNs I indeed do need to understand what those particular failure modes are, don't I?

MR. KURITZKY: Right.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: And you'll have to construct it. If you have more than one failure mode, you obviously have to construct more than one BBN.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: But in doing a proof of concept, we need to pick a failure mode.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: You want to pick one that makes sense, that would really show up somewhere, and then --

CHAIRMAN STETKAR: And we aren't going to see that until you actually invoke the case study, is that right?

MR. KURITZKY: Right.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: And one thing I forgot to mention was, that this report documents the selection of the methods and how we would apply them to the case study, and there will be subsequent reports that will actually document the case studies.

CHAIRMAN STETKAR: Okay, thanks.

MR. YUE: One more thing from this slide is, this table is called a node probability table. This is basically condition of probability. We can perform inference space on this.

People have recognized that in order to have quality or reliable software, some kind of standards or guidance have to be following -- have to be followed during the development process.

There are different standards -- or different standards or guidelines for different industries.

Two examples here, one is Avionic standard DC-178B and this one is -- for example, FAA applies this standard to determine whether the software performs reliably in the airborne system.

The second examples is NUREG-0800, the BTP 14. NRC staff uses this one to determine the -- to assess the software life cycle of development activities for the digital I&C system.

It is pretty natural to capture the quantitative assessment, to -- when we are trying to evaluate a software failure probability, and the BBN model can provide such a function.

So, by using the BBN model, we can -- basically, each node can represent the quality assessment of certain -- for a certain standard or guidelines.

The results of quality assessment can be used in two ways from the literature we have reviewed. The first one is people, they use the quality assessment to determine the insertion, the detection and also, the removal of faults, and eventually, they can determine the number of residual faults in the

software.

There are two references here. For the second, the people, they sometimes -- they just use quality assessment to determine software reliability or failure probability by using expert elicitation.

DR. HECHT: Meng?

MR. YUE: Yes.

DR. HECHT: I just want to make the point, I know a lot about DO-178B, because I worked for five years with the people writing the DO-178C standard, and they will tell you, they will tell you specifically, that basically, there are 66 process objectives, and the way that the standard works is, as you move up the, what are called levels of assurance, D through A, you add more and more of these processes.

So, for example, in moving from level D to level C, you add the branch testing. You're moving from level C to level B, you add decision testing, and then to move from B to A, you add MCDC, which is, I'm not even going to try to describe that. It's a kind of multiple condition decision coverage.

But no one there will say that as you move from, for example, decision testing to MCDC, that you move from 10 to the minus seventh per hour to 10 to the minus ninth per hour, even though that's what you

might imply from the standard.

In other words, they just say that this is what we feel we need to do to move from a flight essential to a flight critical system, and that's, by the way, one of the things that I see as the weakness of this.

CHAIRMAN STETKAR: Myron, has -- it's fine, this is a Subcommittee meeting.

You mentioned a couple of numbers. Has anyone tried to go back and benchmark that type of process, or call it a calibration? You know, you say, "Well, you know, going from," whatever jargon you used, you know, step four to three, or something like that --

DR. HECHT: B to A, yes.

CHAIRMAN STETKAR: B to A, whatever.

DR. HECHT: Yes.

CHAIRMAN STETKAR: Gives people subjective confidence that you can prove the reliability by a factor of let's say, 100. Has anybody gone back to --

DR. HECHT: Not to my knowledge.

CHAIRMAN STETKAR: -- benchmark that, and say --

DR. HECHT: Not to my knowledge.

CHAIRMAN STETKAR: -- indeed, we're

achieving that improvement.

DR. HECHT: Not to my knowledge.

CHAIRMAN STETKAR: Okay.

DR. HECHT: The standard is very useful, because it allows regulators to more or less --

CHAIRMAN STETKAR: Understand that --

DR. HECHT: Yes.

CHAIRMAN STETKAR: -- that objective. I was just curious whether anybody had kind of closed the loop on it, okay.

DR. HECHT: That's what we're all looking for.

CHAIRMAN STETKAR: Thank you, just curious.

MR. KURITZKY: But just to follow on that, that's one of the biggest limitations we have with this approach, also, is that converting from qualitative to quantitative, we don't --

CHAIRMAN STETKAR: Right.

MR. KURITZKY: -- without the evidence, to make the points, we can rely on experts, but it's going to be, as our colleagues from London had mentioned, you have to assume you have a very weak prior, because there is not going to be, you know, any hard evidence to really justify the numbers that your

experts are going to give you.

CHAIRMAN STETKAR: Yes, yes, no, I was just curious whether there was some evidence from --

MR. KURITZKY: From the other industries.

CHAIRMAN STETKAR: Yes, since Myron explained that standard, whether there was some evidence from that industry, to provide some of that anchor.

MR. YUE: I certainly agree. I guess for BBN, one of the assumptions is, if you can perform the development activities following a high quality standard, you have better chance of being more reliable. I guess that's one of the assumptions, underlying the BBN model, here.

Hopefully, we are trying to capture this, using BBN model.

I certainly agree, sufficient condition is not necessarily. Basically, if you are -- if you want to be this reliable, you have to do this kind of thing, but it doesn't guarantee after you do this, you will have this reliability. That's true.

So, basically, after collecting information about a specific software development, we can compare this evidence against the BBN model, then perform inference, to come with prior distribution of

software probability.

The idea will be, as Alan just mentioned, we are going to have a -- probably are going to have a prior developer by BBN model, then use testing data or operational profile, or the operational data, to perform the update of this.

This the work -- this is the first way to use quality assessment. People tend to come with a number of residual faults in software, and we need to convert. Since the ultimate goal is to come up with the software reliability or failure probability, we need to convert these full numbers into the failure probability.

This is called indirect conversion, that's basically from qualitative assessment to full number, then to failure probability.

From the literature we have reviewed, we have seen two kinds of methods. The one is called -- one that's used, size distribution of faults.

Basically, this data collected from the similar software development projects, from the same vendor, that's very software -- or vendor specific data, or in some literature, they are using this so-called fault-exposure ratio.

This is basically obtained from all kinds

of software projects. We have no idea what those projects are, and basically, it's going to integrate many factors that might affect the software reliability.

For example, the software complexity doesn't have a role in this fault exposure ratio. Basically, you cannot tell the difference, but with this fault number and this fault exposure ratio, or size distribution of faults, we can -- we might be able to come up with the software failure probability.

As we just mentioned, the vendor specific data, or software specific data are apparently more applicable, compared to the generic fault exposure ratio, which is from a spectrum of software projects that might have different quality from different industries, for different purposes.

CHAIRMAN STETKAR: Meng?

MR. YUE: Yes.

CHAIRMAN STETKAR: Again, going back, I'm kind of a simple minded guy. So, going back to my ATM example of either eating my card or spewing out more money than I can handle.

In your review of the application of these methods, I see size distribution of faults, or estimated number of residual faults in the software,

without any specificity regarding what the effects from those faults might be.

So, this is just simply -- it seems to be just simply the software doesn't sort of, kind of work the way I sort of, kind of expected it to work, without being very specific about which particular failure modes might be more vulnerable, if I can use that term. Is that true, or this is --

MR. YUE: I think you are right.

CHAIRMAN STETKAR: Okay.

MR. YUE: I think you are right.

CHAIRMAN STETKAR: How do we get from that, though, now, to a prior distribution for three different -

MR. KURITZKY: Failure modes.

CHAIRMAN STETKAR: -- failure modes, that I'm interested for my simple ATM example?

MR. KURITZKY: Well, again, that goes back to one of the points you made earlier, was that -- or that Louis made, was that most of the methods our there don't distinguish between failure mode. They just calculate a failure probability of the software for that system, without making the distinction.

CHAIRMAN STETKAR: Okay, when you say 'most' --

MR. KURITZKY: Right, I don't know.

CHAIRMAN STETKAR: -- that gets my interest, because if you said 'all', that would be exclusive. When you say 'most', it infers that there might be some that do, and --

MR. KURITZKY: Well, let me then clarify. Of all the ones we looked at, Louis, were there any of them that distinguished failure mode?

MR. CHU: Maybe Sergio can explain. I think in your work, there was some discussion on how you partition failure mode into -- failure rate into failure mode, or failure probability into failure mode.

DR. GUARRO: Well, it has to do with, you know, the issue of testing. You know, in the end, and I am expressing my modest opinion, here, that's all, I think that whatever probability value that you use for software, has to be based, or has to give strong weight to the type of testing that you do, for the type of failure modes that you're concerned about.

You know, what I also believe is that before you do that, you have to do the type of modeling that allows you to understand what type of failure modes you're worried about, and then you can link -- you try to test for those, because you know,

in fact, John, you kind of, I would say, anticipated a question I was going to ask, with respect to the issue of the prior, because -- and that's just a word of caution.

I think you have to be careful, you know, if you -- I think a BBN approach, indeed, allows you to combine, you know, expert opinion with actual data.

But, you know, in general, when you do a prior estimation based on things like quality and so forth, you cannot get an average estimation, right.

This is my black box software, and this is, I think, where the failure rate or failure probability, you know, range is.

Well, now, you're trying to translate that into failure probability for specific failure modes, and there are -- you know, you can do it in two ways.

You know, you can spread it around and -- you know, what I'm saying is that you have to be careful not to apply a part of that prior to a failure mode that is not included at all, in the mental model that you used for the prior.

CHAIRMAN STETKAR: I just -- you know, for the purpose of this presentation and this meeting, I just wanted to make sure I understood what was being said on this slide, and I think I do.

I think, you know, in -- as you move forward in this research and go to the case study, it might be useful to examine a couple of different failure modes, rather than just, does not trip the reactor when it's supposed to trip the reactor, if that's one, only to evaluate the types of analyses that need to be done, to bridge the gap between this level of qualitative information, perhaps informed by some, you know, generic vendor experience, to the more specific prior distributions that I think we've just been discussing in the last five minutes, and those are, you know, what -- given this information, what would the prior distribution be for failure mode x versus failure mode y, and what type of thought process do we need to invoke, to do that?

So, that's just -- you know, I know, we'll have more discussion about the bits and pieces, as we go forward through today, but if, indeed, the case study scope is focusing only on a specific failure mode, that might not -- that might not capture how you use this broader sense of information to really -- you know, whether it's parsing it up or however you do it, to develop, you know, different informed priors for those different failure modes.

So, it may be something you want to think

about.

MR. YUE: Okay.

CHAIRMAN STETKAR: Thank you.

MR. YUE: This is a second way of using the quality assessment to come up with failure probability, directly using expert elicitation.

This is one of the examples from this reference, Gran's paper in 2000.

Basically, in the high level BBN model, it's -- it has three qualitative notes, namely, quality of product, solution complexity and the quality of analyses.

Look at this table, each of the three nodes. The first node, quality of product has two states, good and bad. Solution complexity has three states, low, medium and high, and the quality of analyses, it has two states, good and bad, and looking at the combination of all three qualities, and through expert elicitation, they come up with this claim about the software failure probability, with the confidence level.

For example, the first -- in the first row, if you have a good quality of product, low solution complexity and a good quality of analysis, the chance of having the failure -- software failure

probability less than 10 to minus six, is basically 0.01. That's basically what this table tells you.

But our question is, there is -- we don't see any basis for this kind of claim to come up with NPT table like this.

CHAIRMAN STETKAR: And I think that, you know, what -- if you talk to software developers for nuclear power plant safety applications, they will all emphasis that the quality of the product is excellent, that it's a very, very simple software and that obviously, the analyses that are done on it are better than good.

So, that you -- you know, if this is a universe of possible qualities, it's difficult for me to understand how an independent expert can assess this whole universe, regardless of the mapping -- you know, regardless of the left-hand column, what those numbers might be.

DR. HECHT: Thank you, John, you made the point.

CHAIRMAN STETKAR: You're welcome.

MEMBER SIEBER: And the reason for that is, no basis.

CHAIRMAN STETKAR: Yes, well, there's certainly no basis for the numbers in the left-hand

column.

MEMBER SIEBER: Right.

CHAIRMAN STETKAR: But the question is, even -- regardless of what those numbers in the left-hand column might be, that's sort of the -- you know, what's the viability of those anchor points, that we've been looking for, is it reasonable to say that, you know, any of the systems that we'll be evaluating for a nuclear power plant PRA application, would fit in the lower right-hand corner of that matrix.

You know, the quality of the product is bad, it's really, really complicated and nobody has actually looked at it from an analytical perspective.

MR. KURITZKY: Well, I think I --

CHAIRMAN STETKAR: So, you know, this is a nice concept, in -- but --

MR. KURITZKY: Okay, so, let me just make a couple of comments about that.

This, of course, was something that was done, I think under the Halden reactor project in collaboration with -- I don't know who it was, was it VTT, maybe, or somebody?

But in any case, this doesn't apply to our case.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: As Louis was pointing out to you before, what we would be looking at is, you know, like you said, everything would be in that first column.

CHAIRMAN STETKAR: Right.

MR. KURITZKY: Okay, but in fact, what it would be is, it would be -- there would be, you know, further refinement among that first column. We would have multiple --

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: -- theoretically, multiple columns, but they would -- the very bottom -- the most -- the worst one would be probably the top of that table.

CHAIRMAN STETKAR: Okay, okay.

MR. KURITZKY: -- and then we would have some combination. That was the idea of --

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: -- what Louis was mentioning. We would try to construct questions that could help parse out --

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: -- the degree to which they were exceeding the minimum accepted.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: You know, the accepted.

CHAIRMAN STETKAR: Okay, thanks, that helps.

MR. KURITZKY: But unfortunately, the reality is, just like a point I tried to make at the end of my presentation was, when we get to the uncertainty, because again, regardless of how we make -- we could make a more applicable table for our applications, but still, we're relying on some experts to put numbers in that table, with no real data to anchor this thing.

And so, that's where it gets to, you know, we can apply an approach, but if the uncertainty in our answers are so great, you know, that it doesn't allow us to make, you know, reasonable judgments, then that's one thing we just -- I'm just concerned about.

I don't know how that is going to play out, but that's something to --

MEMBER BLEY: Well, our consultants, my members, all of you say there is no basis for this table. I don't know that.

I haven't read the Gran report, and well done expert judgment isn't just a guy picking numbers out of his head. It's a guy with deep experience and access to lots of experiments and data that are out

there, who then brings that all forward, as evidence.

I don't have a clue. I didn't read the Gran stuff, and I don't know what you've looked at, but your experts, if you were to do something like this, would have, as the evidence behind their elicitation, and I'd rather not just call it judgment. It's bringing forth all -- whatever evidence they have, and not -- I'm just curious.

In the -- the Gran papers, or reports, did you get into those deep enough to know whether they had evidence behind this stuff, or they're just arbitrarily laying out a --

MR. CHU: It did have very little information about the experts that they used.

MEMBER BLEY: Yes.

MR. CHU: So, in a way, it's our speculation --

MEMBER BLEY: Couldn't tell?

MR. CHU: Right, it's our speculation that, you know, this table really no -- has no strong basis.

MEMBER BLEY: Okay.

MR. CHU: Because it's hard for people, there may be experts who are familiar with software development and a lot of experience. Even with those

people, they probably won't feel comfortable in giving you a distribution or failure rate or failure probability.

But on the other hand, I would try to defend this method. Isn't this the kind of table that you see in the human reliability analysis?

The human reliability analysis, a lot of numbers. There is no strong basis, also, and we've been doing it for 30 years, as well, and kind of happy with it.

So, for software reliability, we are only at the beginning. We have to start doing something.

CHAIRMAN STETKAR: Could say some methods, then.

MEMBER BLEY: Some methods use this type of thought process.

CHAIRMAN STETKAR: All right.

MEMBER BLEY: But just an aside, I just saw a piece that we got circulated, that staff has now been given direction to go out and come up with a consistent approach to elicitation throughout the agency, which is about time.

MR. KURITZKY: Right.

MEMBER BLEY: And the better work that's been done in this area over the last 10 years,

1 certainly has moved away from what's off the top of
2 the guy's head to what's the evidence behind what
3 they're bringing forward.
4 MR. KURITZKY: Right.
5 MEMBER BLEY: That's why I hope you guys
6 are -- if you go ahead with this, that you're linked
7 with that larger process, so that whatever you do is
8 cognizant of where the agency is headed for, this
9 whole approach.
10 CHAIRMAN STETKAR: I was going to ask, you
11 mentioned that they didn't provide much information
12 about the particular experts, and I'm not -- I've also
13 not read the Gran work. Did they provide much
14 information about the process that they used, to
15 elicit the expert input?
16 MR. YUE: I think it's the --
17 CHAIRMAN STETKAR: Because if they neither
18 provide any information about the experts themselves,
19 nor the process that they used to elicit those expert
20 opinions, then it's very, very difficult to understand
21 what level of confidence you might in these particular
22 --
23 MEMBER BLEY: And it's not or, it's both.
24 CHAIRMAN STETKAR: That's right, it's
25 both. You need to have both qualified experts and

invoke --

MEMBER BLEY: Yes, an expert with a poor process doesn't put --

CHAIRMAN STETKAR: Doesn't do the -- yes, that's right.

MEMBER BLEY: -- pull out useful information.

CHAIRMAN STETKAR: Yes, and the best process in the world without the experts --

MR. YUE: They do have some data about the process.

CHAIRMAN STETKAR: About the process, right.

MR. YUE: Right.

CHAIRMAN STETKAR: Okay.

MR. YUE: Some data about the process. How they come up with, for example, prior distribution, they have some --

CHAIRMAN STETKAR: Okay.

MR. YUE: -- some ways through expert elicitation to do that.

CHAIRMAN STETKAR: Well, but even in the process of how they --

MR. YUE: Select experts?

CHAIRMAN STETKAR: No, well, part of it is

the selection, but in terms of the process that's used to calibrate the experts to develop levels of confidence among those experts, and kind of test those levels of confidence.

As Dennis mentioned, there has been a reasonable amount of work done in that area, over the last, I don't know, probably 10 years, five or six or seven years, within this agency, and there will be more done, you know, going forward here.

Whatever process, you know, you folks use in the case study, should build on that.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: And if there isn't documentation the Gran work, I agree with you, that it's really difficult to have much confidence in, you know, this particular matrix, because you just don't know enough. I mean, it might be really good.

MR. KURITZKY: Right, we just don't know.

CHAIRMAN STETKAR: You just don't know.

MR. KURITZKY: But also, to get back to the idea that since -- unfortunately, in this particular area, in the software reliability area, we don't have a lot of hard evidence that we can use, you know, when -- that Dr. Bley talked about, you know, we don't know, this may be -- actually, have a lot of

basis, if they had -- that these experts bring to the table, a lot of knowledge, they're operational experience and testing and experimentation.

But the reality is, I think with software reliability, we just don't have -- there is just not going to be very much, in the way of hard evidence for the experts to bring to the table.

They will bring their experience. They will bring a lot of anecdotal evidence, I'm sure, but --

CHAIRMAN STETKAR: Right.

MR. KURITZKY: -- to the extent that there is going to be some hard numbers, and some anchor points, that's where we're kind of going to be a little bit shy, and so, that is where we're going to have a lot of uncertainty on the outcome, and that's why we're really, in our approach, we're relying on this to give us kind of a broad prior and it's really the test results that we're going to -- hopefully, will drive the posterior distribution.

MEMBER BLEY: Well, I'd certainly go with that. I mean, you look at the details in this chart, it's -- I mean, if these are priors, they're highly informed priors, each one of them, and it surprises me how tight these distributions are.

I mean, they're decades, but still, they're pretty tight, especially when you get -- they've got embedded assumptions, clearly.

If you leave off the best side, pretty soon they're saying you're very likely to fail, but I even wonder if that's consistent with experience, you know.

MR. KURITZKY: Right, I know, I also have not read that work, so, I don't know exactly what went into it.

MEMBER BLEY: Okay, that's probably enough.

CHAIRMAN STETKAR: Before we move to -- do you have something?

DR. HECHT: Yes, I just want to respond to your comment about, well, this is how it works in human probability, and I just wanted to point out two differences between the human probability or human error probability estimates, and what's being done here.

Number one, those probability estimates, at least it's my understanding as to how they're done, they're quite close to the task.

In other words, you won't ask a satellite operator, you won't use data from a satellite operator

or from an aircraft pilot to infer your air probabilities of a nuclear plant operator.

And number two, software is developed by a group, usually, not by an individual.

MEMBER SIEBER: Yes, it's pieces.

DR. HECHT: So, it's -- I'm saying that the group causes things to change.

MEMBER SIEBER: Well, and it's all pieces, too.

DR. HECHT: Yes.

MEMBER SIEBER: And the integration of that makes a big difference.

DR. HECHT: Yes, it's groups for each piece, and it's certainly groups across, when you do the integration.

MEMBER SIEBER: That's right.

DR. GUARRO: And another way of saying that is that in a piece of software, there are a lot of different tasks.

DR. HECHT: Yes, okay, yes.

DR. GUARRO: Included.

MR. YUE: That's one of the reasons we think we should follow -- we should compare the software development activities against some kind of standard or guideline.

Hopefully, that will make it -- I mean, more attractable.

DR. HECHT: It's still hard.

MR. YUE: Yes, it's going to be hard.

CHAIRMAN STETKAR: If it was easy, we wouldn't be having these meetings.

MR. KURITZKY: Right, I think this is actually a convenient point to break.

CHAIRMAN STETKAR: It is, because the next, at least four slides, are integrally linked to one another.

So, I think what we'll do now is recess for lunch, and I'm going to be generous, because I don't care what time you people go home. We'll break until one o'clock.

(Whereupon, the above-entitled matter went off the record at 11:55 a.m. and resumed at 1:02 p.m.)

CHAIRMAN STETKAR: Okay, let's reconvene and hear more about issues and limitations of Bayesian belief network methods.

MR. YUE: Okay, before we can claim success of BBN -- of using BBN method, we have to deal with issues and limitations of BBN method.

In next four slides, we have issues and limitations about BBN method here.

The first one, of course, is -- we have been talking about this many times, how to relate qualitative information to quantitative information.

Basically, there is a lack of data for this kind of purpose in either low level or high level BBN modeling.

An example for the low level relation is how to relate V&V quality to the detection probability of residual faults, in a relatively high level, as we just -- in the previous slide, the table, the NPT table, for converting the quality assessment to the failure probability. We don't have that kind of data, either.

One thing we are proposing is, benchmarking can be performed to evaluate the software reliability. You can use the collector -- operating data, or applying statistical tests at a low level or high level of BBN modeling.

CHAIRMAN STETKAR: Let me stop you there, because I think it's somewhat relevant to the question that I asked Myron, just before the lunch break.

You say benchmarking can be undertaken to validate the software reliability assessments.

Is that a wish, or you have confidence, because at least in the aerospace industry, I don't

get the sense that anyone has done any of that. Why they haven't, I have no idea, nor maybe is it relevant.

But if that's only a wish --

MR. KURITZKY: I know, it's not only -- I'll let Meng respond more, but it's not a wish. It's something we've planned -- we would like to try and do it --

CHAIRMAN STETKAR: I understand that you'd like to try to do it, but have you thought enough about it to say that if I was to ask you today, at the -- on the pain of death, are you 90 percent confident that you can do, or 10 percent confident that you can do it --

MR. KURITZKY: Since it's a pain of death question, I'm going to let Meng or Louis answer it.

CHAIRMAN STETKAR: An appropriate project manager's answer.

MR. CHU: I think the basic difficulty is lack of data.

So, benchmarking, we call it benchmarking, basically, we need more data, either to validate the model or find out weakness of model, and you modify the model.

So, in that sense, the benchmarking would

help.

CHAIRMAN STETKAR: But there is --

MR. KURITZKY: Benchmarking would help, but what --

CHAIRMAN STETKAR: But this is in the context of informing the priors, not updating, right?

MR. KURITZKY: Right, right. But the point is, could we -- right now, if we had unlimited resources, today, could we go forward and do we think we could do a successful job, at undertaking the bench-mark?

Is there enough --

MR. YUE: Personally, I think it's not impossible because taking at look at operational data, like the French, their claim, they have been using reactive protection systems, the digital ones for 20+ some years, they have 70 of the -- probably, 70 of the nuclear power plants, and they have -- if you look at the operating experience they can claim, they should have quite some data in this -- in terms of operating experience.

CHAIRMAN STETKAR: I understand the claims, and I've heard those claims, and indeed, I've asked people in Europe, whether they can back up those claims with any data, and so far, I've come up empty.

So, what I'm asking you is, is if this an integral part of the research program, you know, what confidence do we have that that second bullet can be accomplished in practice and in real time, meaning, some time, I'm assuming in the next year or so?

MR. YUE: It certainly will not be done in this project. We won't be able to do the benchmarking in this project.

CHAIRMAN STETKAR: But don't you need to do that, as a demonstration of the capability for the case study that you're proposing?

MR. KURITZKY: The benchmarking isn't necessarily a requirement for doing the case study, but I think there was -- to some extent, there -- we're going to use the statistical testing, after the -- you know, at the part -- the statistical testing that we're going to use to come up with the posterior distribution, that, in essence, is kind of serving the benchmarking function, in some regards, for our study, to see whether or not the results we get out the statistical testing would confirm what we have --

CHAIRMAN STETKAR: But be careful, in a truly Bayesian sense, you don't use the evidence that you use for updating a prior as information for developing the prior, which sounds like what you're --

you're saying.

MR. KURITZKY: Not information. We're developing the prior, but to try and give us kind of a feel for how well we think we can come up with the prior.

In other words, if the prior shows, and correct me if I go astray, Louis, but if the prior shows, I'm going to make something up, that the failure probability is 10 to the minus four, okay, as a mean value, let's say, and the test data shows it to be 10 to the minus two.

Clearly, there was something wrong with that prior, I think, and in fact, that after you have enough test data, of course, that prior will disappear on posterior distribution, but at least we know that starting off, the posterior -- or the prior wasn't very accurate.

Whereas, if the prior shows a mean value of 10 to the minus four, and then the data comes up to something in that ballpark, then we say, "Okay, well, that prior probably did a fairly decent job."

Again, it's a single -- it's going to be like a single evidence point. It's not going to be, you know, doing a whole coordinated and regimented study to demonstrate it, but it will give us some

level of increased confidence, if the data from the statistical testing is in the ballpark of what we have come up --

CHAIRMAN STETKAR: I'm trying to understand how we get to the ballpark. I don't care about how you collect the information, at this point, to update that ballpark.

We've already established that people will do some sort of evaluation, qualitative evaluation of attributes of the software development process, as input to this -- the Bayesian belief network, and somehow, in the light of the matrix that you showed on the preceding slide, there needs to be some way of, let me call it calibrating, anchor points for that qualitative assessment, in the analogy that Louis used for human reliability.

You know, various methods, and I don't want to go into HRA methodology, but various methods have been proposed of trying to -- you know, how do you establish those anchor points, so that the relative evaluations are, indeed, anchored, in terms of an absolute numerical perspective?

That's my interpretation of this second bullet on this slide, and maybe I'm not understanding the second bullet, quite right.

MR. KURITZKY: Well, again, I mean, Louis, you can step -- do you want --

MR. CHU: I think the second bullet is not what we plan to do within the current project benchmarking.

The only exception is what Alan just explained, in the sense that we use BBN to come up with a prior distribution, and then we do statistical testing.

CHAIRMAN STETKAR: Okay, but let me stop you there. How do you use BBN to come up with a prior distribution?

I can understand in qualitative terms, relative goodness and badness of various attributes, and I don't care what those attributes are, right now, but I understand that.

How does one actually translate that degree of goodness and badness into numerical estimates, whether they be 10 to the minus two, as a mean value, with a range of a factor of, you know, 50, or whether they be 10 to the minus four, as a mean value with a range of 1,000.

MR. YUE: I think this slide, this slide basically, as Louis said, we can use BBN quality assessment to -- two ways of using quality assessment,

one to come up with the residual number of faults inside the software.

Then find some kind of vendor-specific data, or even generic data, to relate the number of faults to the failure probability, considering operational profile.

CHAIRMAN STETKAR: I hear you saying those words, and what I'm asking you is, the next step beyond saying those words is, is there confidence that those words can actually be implemented, and I hear you saying that that may not be part of this project, whatever this project is --

MR. KURITZKY: Let me just clarify -

CHAIRMAN STETKAR: -- it strikes me as it ought to be part of the case study, because the case study, as I understand it, is sort of a demonstration of the methodology, and not just -- let's just assume that this is 10 to the minus six, and let's just assume that this is 10 to the minus two, and showed that, indeed, we can do a Bayesian update.

We all know that we can do Bayesian updates.

MR. KURITZKY: Right, let me just clarify what's part of the project and what's not.

The benchmarking, where if we come up

with, for instance, a distribution table, similar to what was done for the -- in the Gran example, but of course, you know, for our own application, okay.

Whether or not those values that are in that table, you know, because we have limited evidence, or maybe none, to be able to come up with those values, it's going to be based on expert elicitation, and you'll hear more about that shortly.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: And it's going to be a lot of -- well, you know, as Dr. Bley said, it's not just people, just pulling numbers out of the air, but they're bringing their experiences and their -- whether it's anecdotal or actual experience and exposure to different tests and experiments, to bear, to come up with some estimates.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: Okay, what we want to do is have benchmarking, where we try to look at actual software, see where it would fall in our matrix, have actual operating experience or test data for that software, and then see whether it comes reasonably -- in reasonable agreement with what our matrix would show us.

Now, that is what's outside the scope of

the study. What's within the scope of the study is going through the expert elicitation, and using one of these two approaches to convert the qualitative information to quantitative values, and that's where we have either the direct conversion or the indirect conversion, and you saw in the example with Gran, that was a direct conversion.

But either way, we're going to use what -- we are going to use one of these approaches in the case study, to come up with values, but what's not in the case study is to bench-mark those values against actual other software, to determine whether that was a good -- whether that was a reasonable -- you know, came up with reasonable results, except to the extent that we will use our test results, for our particular software.

We know where our software would fall in our matrix, and we use that test result to say, "Okay, we did it. Was it reasonably in agree or not?"

That's the only thing that we're doing, in that regard.

MEMBER BLEY: Yes, and I think one important thing there is, you use some of the tests, once you get data, to do your update, use some of the tests for Bayesian inference, to see that the data --

that your prior isn't too heavily informed, you know.

In simple language, if your results pile up at one end of your prior, if your posterior is piled up at one end of your prior, you know there is something funny with your prior. There is a lot better explanations than that.

But as a minimum, you need to do that, and if it turns out they are doing that, then you've got to go back and figure out what in your process led you to overly constrict the prior? What was it? Was it information, and why wasn't that information applicable to what you're doing now?

So, there are a number of actual tests for doing that, and you've got to do something.

CHAIRMAN STETKAR: Thanks, I was just -- I now understand. I was hanging up on that second bullet, as a way of, you know, some way of numerically calibrating what I better understand, now, what those priors will entail. Thanks.

DR. HECHT: With respect to your explanation on slide six, can you go back to that for just a second?

You said basically, the number of faults that are in the software and basically, the probability that a given fault is going to be

translated to a failure, I guess my problem with that is first of all, how many faults do you know remain in the software?

And number two, how do you measure fault exposures, and specifically, in your work, in your past work, you've used software reliability as a rate, which is something I'm familiar with and understand.

So, you Markoff models, you have a failure rate. So, going back to the French example, if we have 70 reactors years of digital feedwater control experience, let's just say, and we have three failures, you can impute some kind of a failure rate.

But with respect to what I think you're trying to do now, which you call discrete, in other words, the probability of failure on demand, or the probability of success on demand, how do you know the number of demands that a particular application has been exposed to, so that you can calculate the defect exposure ratio, even if you are able to know what that -- those number of defects were?

MR. YUE: The first question, number of defects, as the previous slides say, you can use quality assessment, as a factor, because it's going to affect the insertion, the detection and the removal of faults, during the software development process.

For each requirement specification design, coding, testing, you'll have these different phases. You might introduce fault and also, you might be able to detect and remove some of the faults.

Based on some software project development activities, some data, we are trying to model this mechanism to come up with basically, an estimate of number of faults, eventually in the software, when it is delivered.

DR. HECHT: Well, in that digital feedwater control study you did from Idaho, do you know how many defects remain in that DFCS software?

MR. KURITZKY: Now, you're talking about the ATR -- the system that we're going to use for this -- the case studies, or the previous --

DR. HECHT: The previous.

MR. YUE: The DFWCS.

DR. HECHT: Yes.

MR. YUE: We didn't look at software.

DR. HECHT: Well, you have the software failure rate, as part of the Markoff model, I think.

MR. KURITZKY: Right.

MR. YUE: Right, that was a place that just --

CHAIRMAN STETKAR: They just threw a

number in there for it.

MR. KURITZKY: Right, that was just a place holder. We didn't actually do any analysis for that.

DR. HECHT: Okay, well, then let's take the example that you're going to use now.

How will you estimate the number of residual defects in that software?

MR. YUE: In that sense, we need to contact -- I mean, we need help from those development teams, who developed the software for this. Basically, ask them to look at each phase, how many faults they detected, and they corrected, and because we will never know exactly how many faults are over there.

So, basically, we are going to try to estimate the maximum number of faults for each -- at each phase.

So, number of faults from the previous phase will be carried over to the next phase. Basically, from that point, you can aide the faults left in the previous phase, to the current phase, and when you go through the software development life cycle, you should be able to come up with estimate of number of faults left in your software.

DR. HECHT: I'm not quite sure I understand that. I would point you to some work that was done by IBM, using a model called SWEEP and STEER -- no, SWEEP.

MR. KURITZKY: S-W-E-E-P?

DR. HECHT: Yes, and I will send you the name, as soon as I come up with it, where they estimated number of defects in the original software.

It's very controversial. I'm not sure I believe it, but that is one method of doing it.

Basically, they have a -- they use a distribution called a Rayleigh distribution, so, at the beginning of code and unit tests, from the time distribution that you see of failures being generated, you can draw a curve.

CHAIRMAN STETKAR: Myron?

DR. HECHT: Yes.

CHAIRMAN STETKAR: Let me interrupt here. I don't think that's the purpose of this meeting, to get into a lot of details about precisely how one expert or another expert might estimate these various inputs.

DR. HECHT: Okay.

CHAIRMAN STETKAR: So, it's a useful discussion, but probably too detailed for what the

purpose of this meeting is.

I think we're trying to get an understanding of the basic process that they're going to use, what they're going to do, rather than very, very detailed things about precisely, how they're going to do it.

So, you know, references to methods that others have used, I think are relevant, you know, spending a lot of time on discussions of specific details may not be productive, you know, unless we want to stay until about eight o'clock tonight.

MR. KURITZKY: But let me -- you're exactly right, for the purpose of this briefing, we don't, but for the purpose of our project, we're very interested in that.

CHAIRMAN STETKAR: That's right. I thought it's certainly relevant information. I just want to make sure that we have a chance of --

MR. KURITZKY: Right, we could talk about --

CHAIRMAN STETKAR: -- getting through all of the, you know, all of the parts of the presentation today, in some, you know, reasonable sense of time.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: And with that, Sergio

is going to add some more detail.

DR. GUARRO: I am going to move to something very general.

It's a question regarding the use of the expert elicitation, and have you thought specifically how you're going to address the question of design errors?

In other words, not -- you know, I know that you want to take into account the quality of the process, et cetera, et cetera, and ask the experts to assess some level of quality that you may, you know, try to translate into a range of likelihood for a prior estimate, or something like that, if that's my understanding of what you intend to do, is that correct?

MR. KURITZKY: Yes.

DR. GUARRO: Okay, so, within that framework, how do you plan to make sure that the expert, when he or she thinks of, you know, quality and likelihood of software failure, as in their mind also, the specification process, you know, the -- are you translating the concept of operation, or whatever the software has to do, into a correct specification, correct, meaning that suitable to face the gamut of realities that the particular logic has to counter, so

to speak?

Obviously, this refers to a broader, you know, range of application than just the simple shutdown system, because there, you can argue that, you know, the function is so clearly defined that there is not a lot of ambiguity about the design.

But in the context of what John was saying before, you know, in the future, you might have something that could control the rate -- a rate of, you know, safety injection to match the decay heat, or whatever, something more complicated.

CHAIRMAN STETKAR: Well, but part of that, we discussed before lunch, the notion of -- if I understand what you're saying, Sergio, is if in any type of expert elicitation process, you'd have to be very careful to ensure that the experts understand the scope of the problem, and that they, indeed --

DR. GUARRO: Yes, that's what I'm thinking about --

CHAIRMAN STETKAR: -- and you run some tests on those experts, to understand -- to make sure that they understand the scope of the problems, so that, you know, they are aware of the kind of issues that you bring up.

DR. GUARRO: I'm saying this because of my

familiarity with how the software experts interpret software failure.

They typically say, "I am given a specification," and that's out of the picture.

CHAIRMAN STETKAR: But we've run into the same --

DR. GUARRO: So, I am saying --

CHAIRMAN STETKAR: -- types of problems, with other types of expert elicitation, where design --

DR. GUARRO: And unfortunately --

CHAIRMAN STETKAR: -- hardware designers think of certain --

DR. GUARRO: Unfortunately, you know, in the world in which I operate, in this base arena, and you know, you guys have heard this before from me, the experience, the fact is that a large majority of the failures have come upstream of the specification, not downstream of the specification.

CHAIRMAN STETKAR: Yes.

DR. GUARRO: So, now, the nuclear reality may be different, we don't know.

CHAIRMAN STETKAR: We don't know.

DR. GUARRO: So, that is just -- I'm just saying that, you know, there may be a mind set in the

expert, to think that they have only to answer the question, whether given a specification, the software quality is going to be high.

CHAIRMAN STETKAR: Yes.

DR. GUARRO: Okay, well, you know, I think the point is made.

MR. YUE: Our plan to look at the software development life cycle, basically, specification is one of the steps they are going to look at.

The design arrows, from this point of view, it should be included in the consideration, when we are trying to do the expert elicitation in the future.

CHAIRMAN STETKAR: Okay, that, again, I hope that the planned case study, you know, will, indeed, be a realistic test of some of these concerns.

So, you know, in particular, you know, how is it going to be done, is probably -- you know, it's obvious, there are concerns about that, today, but the primary reason of this briefing is to discuss what is going to be done, and the rationale for why that is what it is. Got that?

MR. KURITZKY: Okay.

CHAIRMAN STETKAR: Now, you can get to the third bullet.

MR. YUE: Well, third bullet, basically, we have discussed, the difficulty in undertaking bench-mark studies should not be under-estimated, specificity and applicability of data to this -- to relating -- to relate qualitative information to quantitative ones.

CHAIRMAN STETKAR: But for our purposes, that's a generic, to be addressed later type issue.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: At least for the near term, the next phase of this project.

MR. KURITZKY: Right.

MR. YUE: Right.

CHAIRMAN STETKAR: Okay.

MR. YUE: And also, Louis just mentioned this idea of regression analysis for Frestimate, it could be extended to directly estimating software failure probability, as long as you have data, but the problem is lack of data, and we have no idea what kind of data they used in the previous software projects. They are generally proprietary.

The second issue with BBN method, complexity of BBN structure, the early software is developed and based on some kind of standard or guideline, like for example, BTP-14.

If we want to develop a model again, based on this kind of standard or guidelines, the BBN model could be very complex, like we just mentioned, we want to evaluate the software development life cycle. We have design specification, design coding, testing, where we use this kind.

For each of them, BTP-14, that indicates like, functional characteristic and process characteristic and for each of them, a bunch of factors have to affect this -- one of these characteristics.

If you create a model based on this, it's going to be pretty complex, and you need to create a large number of NPT tables, as we just showed.

CHAIRMAN STETKAR: Have you thought much about that complexity, because you know, in your -- let me draw the analogy, from my experience.

Back in the old days, when people were doing human reliability analysis, and I hate to draw these parallels, but there are some, you know, the notion was, the only way to do human reliability analysis is that I had to look at each step, in an emergency operating procedure, that might have 100 steps in it, and develop a little model for that step, and that the only way the people could succeed is if

they successfully go through each and every step.

And there were recoveries, and you started to develop these very, very complex models, which we quickly learned, had very little bearing on the likelihood of people making errors.

So, have you thought very much, if you say, "Well, you know, if we just follow the guidance in BTP-14 with all of its various sub-text," have you thought much about that, that sort of experience, that just simply creating a very, very complex model of a set of requirements may not have very much direct relevance to your -- the thing you're trying to achieve?

MR. YUE: That is the second sub-bullet of this slide.

CHAIRMAN STETKAR: Okay.

MR. YUE: Alternatively, it may be more practical to develop a higher level model. We call it -- or you can call it a simplified model.

MR. KURITZKY: But excuse me, just -- but to more directly address Dr. Stetkar's point, we would say alternatively, it may be more practical and possibly more accurate, because the point being that not just would it be more complex, but it might not actually give you --

CHAIRMAN STETKAR: You might actually have more relevant —- although the uncertainty -- the perceived uncertainties might be broader, you might actually have a much better characterization of the problem.

MR. KURITZKY: Right, but we wouldn't -- that's the problem, we're not really going to know that until -- that's -- well, I don't want to cut Meng off, either, and talk about how we were going to look at --

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: --- types of models, but ultimately, we don't have -- we're not having -- even after our project, we're not going to have enough data to rule, you know, decisively on that issue, but we would hope to get some input, as to how relatively accurate the various models are, the more complex versus the less complex.

MR. YUE: Basically, as Alan says, we are planning to do a comparison based on --

CHAIRMAN STETKAR: Yes, you're planning on going both --

MR. YUE: Right.

CHAIRMAN STETKAR: -- both directions?

MR. YUE: Right.

CHAIRMAN STETKAR: Okay.

MR. YUE: And to see how -- whether the simplified version can provide comparable results, before we feel more comfortable to go to the simple model.

CHAIRMAN STETKAR: Well, if the results are different, it doesn't necessarily mean that you -- you feel more comfortable with the more complex model.

MR. YUE: Right.

CHAIRMAN STETKAR: That is kind of what we learned from human reliability, is people were comfortable with those complex models, because by definition, complex models must give us -- must be better.

MR. KURITZKY: And that's the --

CHAIRMAN STETKAR: And indeed, they're not, all the time.

MR. KURITZKY: That was the exact point I was going to make, after you said that, it was based on your comment, up until -- and if we've done the two different models, and they came up with different results, then we would say, okay, you know, because we want to see can we do -- is the simple model going to give us a similar result, then let's go with the simple model.

But if they gave us different results, and you know, it would be okay, well, the simple model doesn't give us the right number, but that's not the correct --

CHAIRMAN STETKAR: But that's the correct mind set.

MR. KURITZKY: Right, that's right.

CHAIRMAN STETKAR: They're different, why are they different?

MR. KURITZKY: Exactly, and it doesn't necessarily mean that the more complex one is correct.

CHAIRMAN STETKAR: Right.

MR. KURITZKY: Exactly right, it's a good point.

MR. YUE: Okay.

CHAIRMAN STETKAR: But the point is, you are planning to apply both --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- modeling, you know --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- methods.

MR. KURITZKY: And a third.

CHAIRMAN STETKAR: Okay, a third, okay.

MR. KURITZKY: Some time.

MR. YUE: On the -- also, in this slide, it's -- we are considering to use failure likelihood index method, the FLI model, from the human reliability analysis, to perform, to create a simple BBN model.

The third issue related to BBN method is causal relationship representation and BBN structure.

Basically, BBN reflects the kind of relationship of different nodes representing the relevant events. If the -- basically, the arrows, they point from parents node to child nodes, that represented influence of parent node, on the child nodes.

If a single child node -- if a single child node has multiple parents nodes, basically, that means you have built a high-dimensional NPT table, when you are trying to quantify the BBN model, which is —- I mean, building NPT table is difficult. Building high-dimensional NPT table is more -- it's even more difficult.

So, from the literature we have reviewed such a structure, it's frequently modified by reversing the direction, that basically, they can -- for example, a bunch of factors that can affect the completeness property and instead of pointing from the

-- these factors to the completeness, they reverse the direction. So, they can only have very low dimension, NPT table, it's like one-by-one. It's not one-by-one, I mean, just like only has one primate over there.

That makes it much easier and however, we don't feel quite comfortable with that, because we don't know -- we know that we present a different -- probably, represent a different physical meaning, so, another point in this project is, we are going to compare two types of models.

One is, the arrows will point to the different direction, and see how different are the results might become -- might come up with.

MR. KURITZKY: Just to be clear, we're not planning to do that for the whole case study. This is just going to be a simple little example, on the side.

MR. YUE: Right.

MR. KURITZKY: We're going to try and see how they compare to each other.

MR. YUE: Right, just as simple examples.

MR. KURITZKY: And then based on that, we'll determine whether or not we can apply it, we think we can apply it to the case study.

MR. YUE: I think we can do this from the -- in the background slide of BBN, you know, you can

do both forward inference and backward inference.

DR. GUARRO: Can you elaborate a little bit about this reversing of the nodes, because you know, I have some familiarity with BBNs, but it seems counterintuitive that if you say that, you know, rain and wind affect how wet you are, you know, in a given day, that you can reverse the nodes and say that how wet you are influences how much wind and rain there is.

MR. KURITZKY: Can you go to that --

MEMBER BLEY: I think influences would probably be the wrong word. It would be conditionally, the likelihood of wind and rain would be different, depending on how wet you were.

So, it's not causal, it's the --

MR. KURITZKY: Yes, I would say --

MEMBER BLEY: It's causal, going the other way. This way, it's inference, going back.

DR. GUARRO: Yes, I understand that part, but --

MR. YUE: So, basically, this simple model here, in general, people tend to think you have wet grass because sprinkler is or rain -- or it is raining.

So, causal relationship is like sprinkler

is on, or it's raining, it's going to cause this wet grass.

DR. GUARRO: So, what you're saying is that you -- if you have a causal correlation -- a causal relationship, a cause-effect relationship, which happens to be too complex, you may revert it into a correlation that goes in the reversal --

MR. YUE: Right, like this one, like this one --

DR. GUARRO: Okay, I understand that.

MR. YUE: Okay.

DR. GUARRO: Thank you.

MR. YUE: But again, as we said, we don't feel confident about whether we can model -- we can change the model like this, only because you are trying to have a simplified NPT table.

In this project, we are planning to use simple examples to compare results.

MEMBER BLEY: I'm sorry, say that one again. I just didn't understand.

MR. YUE: Okay.

MEMBER BLEY: The last sentence you said, I just didn't parse it.

MR. YUE: We want to compare the two different structures, compare their results, and see

how different they can be.

MEMBER BLEY: Okay.

CHAIRMAN STETKAR: As you mentioned, that's a little side step.

MR. KURITZKY: Side, right, exactly.

CHAIRMAN STETKAR: Yes.

MR. YUE: Some other potential limitations example is like unexpected or controversial results may be raised in applying the software evidence to BBN model, and in one of the literatures, we have seen basically, that saying, if you aide more supportive evidence, you might decrease your confidence on the results, basically, it's something like that.

It's counterintuitive, but the argument in that paper was, you have to consider the subtle interplay between different nodes.

In this case, we need to -- we might have to revisit or come back to look at BBN model and see that whether there is some problem with that.

Another issue is hard evidence, usually, hard evidence is very clear, because it represents a realization of a particular state of a node, like it is raining, like it is cloudy or it is not cloudy. These kinds of statement is hard evidence, and soft evidence, some people, they use that.

Soft evidence can be, you have 10 percent of chance of being cloudy today, and a 90 percent chance of being not cloudy today.

This kind of evidence is called soft evidence, and basically, hard evidence will be preferred, because the meaning of soft evidence is not clear in this community.

MEMBER BLEY: Can I take you back to your slide four, again? This kind of resonated with me, and just take a little diversion.

Some years ago, we worked with some physicians that -- to build a little expert system, to look at help with diagnosis, and the kind of thing we saw fits this picture. You know, you could have used this.

But the kind of thing we saw is most people ask the question, "Ghee, if I've got the disease," the physician says, "Well, I think that probably comes from this case or that case," but what we found was, that's not the best judgment the physician can make.

The best judgment they can make is the way you've laid this out, and the causal direction, given I've got this symptom, the sprinkler work, what's the probably that I have wet grass?

So, working in that top-down in your Bayesian network seems to be the way we can best make a good judgment, and then you can use that to go backwards, if you want to, to understood the likelihood of different things contributing.

MR. YUE: Right.

MEMBER BLEY: So, I'm a little -- and your last slide kind of hinted at this, too.

So, I'm a little confused about the statement that sometimes, it's easier to go backwards, and I'm not sure experts work backward as well as they work in a forward direction, because what makes them experts is their knowledge of how things work, which is the forward direction on a chart like this.

So, can you say anything about your idea of why some times, working backwards would be easier, and I think that means easier for the expert to lay out the probabilities.

MR. KURITZKY: Meng, I'll let you speak in a second, but I think the point here is, we agree with you, I think, that it's more natural of the experts to work in that forward direction.

The problem is --

MEMBER BLEY: That's where they probably have the best knowledge.

MR. KURITZKY: Yes, they have the best knowledge. The problem is going in that forward direction, is that we could end up with a tremendous number of node probability tables that have to be populated, with limited evidence, and so, it becomes a fairly -- you know, maybe intractable problem.

Whereas, if you go the other direction, you could greatly limit the amount of nodal probability tables you need, but the problem is, you may have less tables to have to deal with, but you're going to have -- but it's going to be more difficult, you know, maybe less confidence that the experts can give you, the values you need to put in those tables.

MEMBER BLEY: And my worry with it is, it's much more likely, the experts don't think broadly enough to get it right, going in that backwards direction.

They lock in on, you know, the last time I saw the grass was wet was because the sprinklers were working, and to remember, oh, yes, but it could have been rain, is a little difficult.

So, I get -- I suspect -- I mean, it's tough enough to get good judgments from the experts, and if you put them in this boat, because it's less things to evaluate, I think it's just asking -- for

getting back information.

CHAIRMAN STETKAR: It's especially true, if the experts have always lived in the Mojave Desert, for the last 20 years, and that's their realm of experience.

MEMBER BLEY: Then I don't want them there.

DR. GUARRO: My experience would be, what is really difficult is if you have a node in which you have the confluence of several effects, okay, now, to define the joint probability distributions over so many different input variables, that is, you know, from a mechanistic practical point of view, rather difficult. It's just, you know, because it's not just -- I mean, it's relatively to -- easy to say, given that, what is the probability of this other thing.

But if you say, given that and that and that and that and that and that, and you have a list of 20 things, what is the joint probability, you know, of each combination of states of the 20 things, to give you the state of the output, which is really what a BBN does, because if you have five input nodes, each with 10 states, and 10 is many, but you know, I'm just for the sake of expert elicitation, then you have to do the combinatorial of all of those, and each one of

that -- for each one of that, define a conditional probability of a state of the output, and that is --

MEMBER BLEY: I agree, that's not easy. I worry that trying to simplify that by starting at the end and saying, "Where did I come from," is really asking for trouble, where I think I'd rather see you go back to the full model and say, "There are three reasons I don't think these three states really matter much," and I'll drop =---

CHAIRMAN STETKAR: Simplify the number of states.

MR. KURITZKY: I'll purposefully simplify it. The other way is intuitively simplifying it without getting the -- knowing what you're talking about.

DR. GUARRO: I agree with you, I mean, I will attack that probably trying to determine this 20 possible inputs, which ones are really important and which ones may not be as important, I don't know.

CHAIRMAN STETKAR: Well, we're all pretty close.

MR. KURITZKY: But anyhow.

CHAIRMAN STETKAR: But I think you hear the message. I mean, the goal of this is to gain some inference that we have the best input from those

experts, whoever they are, and that we -- the model structure is appropriate, to elicit that input and you know, Dennis' example, there, there apparently are counter-examples, to say that the reverse process may not achieve that.

MR. KURITZKY: Right, and like --

CHAIRMAN STETKAR: Regardless of how simple it may look in -- you know, processing --

MR. KURITZKY: Right, on the paper.

CHAIRMAN STETKAR: -- the numbers, sort of --

MR. KURITZKY: Right, and I think Meng acknowledged that, you know, we had a lot of concerns about going in the other direction, too, but we'll take a look at it.

MR. YUE: I think I'm done with this slide. Uncertainty, treatment of uncertainty. Uncertainties are intrinsic, so, we need to find out the distributions for the nodes, and also, for node probability tables.

Again, we need data or knowledgeable expert for this purpose, to assessing the -- to assess the distribution, and again, the same issue comes back, how to relate the qualitative information to the quantitative information, like how to relate the

software development activities into software failure rates and the probabilities.

We are planning to address the uncertainty of the BBN model in a typical PR -- typical way done in PRA, by using different assumptions or models.

One example is, we just mentioned, we can use a simple model, instead of complex, and do a comparison between them by using a FLI based model.

We've developed this procedure for coming up with a BBN model, and basically, the procedure here, they are more likely to remind ourselves, what we are supposed to do, when we are doing the BBN model.

The first step would be system familiarization. We need information on system design and operation test, operational data, and also, software development activities description, such as debugging data or V&V reports.

We also need to identify expert in this area, in generic software development, and also, we prefer, the example system will be developed based on BTP-14, the NRC guidance.

So, the model is more applicable to U.S. power plants, but this is not always the case. For example, the example system, ATR, Advanced Test

Reactor, they will not develop, according to BTP-14, they will not develop that.

So, in this case, we need to interpret the standard, this ATR system followed, in terms of BTP-14, when we are trying to evaluate this example system against the BBN model.

The major issue could be unavailable information, or documents of the example system. We are having that problem for time being.

Before we create a BBN model, we need to resolve the high level issues here and against conversion of qualitative information into quantitative information. The state-of-art in this area is very weak, and there is a need for additional research, such as benchmarking, we just proposed, but again, it's not going to be done in this project, the benchmarking.

The third one is, we are going to develop high level BBN model. We haven't decided the high level BBN model structure yet. It is -- it could be a direct conversion or indirect conversion, and after we decide how to do it and the high level BBN model will be created.

We need also to develop expert elicitation approaches. Basically, we see two types of expert

elicitation. The first type, we call it Type I expert elicitation, which -- who are familiar with the software development activities and also, they should be familiar with the software development standards and the guidance.

The second type of experts, we call Type II experts. They should be familiar with development activities of a specific software, like the example system software for the ATR we are trying to evaluate in this project.

It is questionable, whether adequate Type I expert can be identified. Basically, we need Type I expert to help us develop the BBN structure, and also, the node probability tables for BBN quantification.

Compared to Type II -- Type I expert, Type II expert should be easier to get, because we always have software developers from a specific project.

Some other issues in the software, in the expert elicitation approaches will be considered. For example, we are trying -- when we design questions for elicitating experts, we will -- how to avoid a systematic bias, that's one of the issues that will be considered in this expert elicitation approach.

Then we can develop a complete BBN model,

basically, create the low level BBN sub-set, by accompanying for specific standards.

For example, the BTP-14 fault is -- project, and again, we just mentioned, we want to have a simplified BBN structure and also, we want to keep the overall BBN structure as simple as possible.

Expert elicitation to quantify node probability table, we will rely on Type I expert for this purpose, and PRA expert will be -- undertake this task.

Collection of evidence, basically, we will rely on Type II experts to collect evidence from a specific software we are trying to evaluate.

The rest of the steps, they are relatively straightforward, performing BBN analysis, there are a bunch of software tools that can be used for this purpose, and we can perform sensitivity analysis to address modeling assumptions and issues, and also, the bench-mark studies, to evaluate a BBN model, that will be outside of this BBN, but we consider this one of the important steps in terms of calibrating or validating the BBN models.

Current status of the example system, we are using these advanced test reactor from Idaho National Lab. They have five so-called in-pile tubes.

They have a loop operation control system. It is a distributed control system. The software is commercially available. It was developed by this company called Metso Automation.

They used these standards, ASME Nuclear Quality Assurance, not BTP-14.

So far, the information we have acquired from Idaho National Lab that can be used to support the BBN development, including the software quality assurance plan, for this new operation control system upgrade. It went through two upgrades, after the loop operation control system was installed.

To us, the most recent one is the relevant, and also, we have some documents about the software V&V plan. Again, it's only a plan. It's not a V&V report.

A configuration management plan, that contends -- that contains some information about software development activities, and also, some details of the previous upgrade of this control system.

We certainly need more information, especially this V&V report, of the most loop operation control system upgrade. We are going to get it, but we don't know how much detail it has in this V&V

report.

Also, we are trying to get more information from NRC staff, because we know that NRC is doing the licensing, we are wondering whether the staff member has some kind of check list to perform the deterministic review, to review the software life cycle development activities.

Those information will be used for when we are trying to develop the questionnaires for experts.

So far, they have realized, it is very difficult to get the information we exactly need.

CHAIRMAN STETKAR: Meng, let me ask you kind of a pragmatic question, here, and I read in the report, the difficulties that you have identifying, you know, something that looks like and talks like a protection system, and for which you could find appropriate documentation, both of the system design and the software development process, so that we could have a system for the case study.

MR. YUE: Right.

CHAIRMAN STETKAR: Recognizing those difficulties, in practice, if I looked forward toward doing this for a real protection system in a real nuclear power plant, does this mean that -- what constraints does that information impose on who can

actually perform these analyses?

You know, is this strictly within the regime of the commercial software developer? Is it within the regime of the licensee, the applicant for a particular -- whether it's a software upgrade, you know, Oconee, for example, or a particular combined license applicant, who might want to use risk-informed information?

Who does these things, and if they do these things, how in practice -- this comes back to NRC staff, does the NRC staff have the resources to be able to make any judgment about the quality of those analyses?

MR. KURITZKY: I think the answer to the first question, in my mind, I envision it's something that really is going to be needed to be done by the licensee in coordination with the vendor.

CHAIRMAN STETKAR: Okay, that's pretty clear.

MR. KURITZKY: So, then I think the only ones who are going to have access, to have the information -- hopefully, they will have access to the necessary information, and I don't even know if that's guaranteed.

CHAIRMAN STETKAR: I was going to -- the

hopefully is the qualifier that --

MR. KURITZKY: Right, sure.

CHAIRMAN STETKAR: -- is a bit of a concern.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: But then it comes to the other question, which is, okay, given that the licensee, the spear-heads, you know, is in charge of the analysis, you know, they'll be getting a lot of support from, let's say, the vendor.

Okay, then how does the NRC, when we get a submittal, how do we go ahead and make ourselves feel comfortable with that -- that what's in there is adequate for, you know, approval?

We haven't --

CHAIRMAN STETKAR: I understand, the research, that isn't necessarily your concern.

MR. KURITZKY: No, but it is, and we wouldn't --

CHAIRMAN STETKAR: But it's a practical --

MR. KURITZKY: -- want the -- it's priority, an approach that has not ability for us to be able to check it.

CHAIRMAN STETKAR: Right.

MR. KURITZKY: So, obviously, a simpler approach that we could also do on our end, to check or audit would be ideal, and at this point, I don't think we can really say.

We're going to go through this approach, believe me, I get a lot of headaches, just thinking how even the licensee or vendor is going to do this, you know.

I get headaches, thinking about how we're going to do it for a simple narrow case study, sweats, when I think about how a vendor is going to do it for an entire, you know, complex system, and --

CHAIRMAN STETKAR: Well, and the reason I raised it, I understand that, and you know, I think we still need to keep in mind, the fact that whatever methods are developed, you know, this is not pure research, that we're trying to develop methods that can be implemented by real risk analysts in real time, for real systems.

Because you know, if we develop methods that can't be implemented practically or have confidence that whatever those products are can be reviewed practically, it's not clear, what we're doing.

MR. KURITZKY: Right, and I think at this

point, we can't make that -- we certainly can't make that call, at this point.

CHAIRMAN STETKAR: Yes.

MR. KURITZKY: So, we're going forward in the path that we think is most appropriate, and as we go forward, I think we have to be very conscious of that.

I mean, we may -- ideally, we go through this exercise, and ideally, we'll say, "Okay, this took a fair amount of effort," but we were able to accomplish is, but now -- and here is ways that we think it can be simplified, going forward, based on what we've learned, and should be more manageable by the licensees, et cetera, and here is some ideas of what the NRC could look for, when they get such a submittal, to have a reasonable assurance that, you know, it works.

That's ideal. Now, how that is going to play out, I don't know, until we get -- really, honestly, have to go more forward through the process, before we can make judgments on that.

CHAIRMAN STETKAR: Have you had any input from industry, on this process, yet, formally?

MR. KURITZKY: Not per se. We did -- I did talk to EPRI. Remember, we have a Memorandum of

Understanding with EPRI.

CHAIRMAN STETKAR: Yes.

MR. KURITZKY: And so, under the MOU, we described a little bit about what we were planning to do. But we haven't had any type of in depth discussion.

CHAIRMAN STETKAR: You're waiting for comments on the draft NUREG, to —-

MR. KURITZKY: Right, and we have --

CHAIRMAN STETKAR: -- to sort of elicit that input?

MR. KURITZKY: That's right, and we have directly talked -- EPRI is a member of -- we actually have an external peer review group. They are a member of it.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: They have two very good people that hopefully, will get a chance to -- many very good people, but there is two particular ones that I think are very knowledgeable in this area, that we're hoping to be able to get comments from.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: So, we do hope to get comments in that regard, but it's out for public review, too.

Last time we sent the NUREG out for public review, we did get information from, for instance, the Westinghouse Owners Group and some other industry, NEI, and some other industry organizations.

So, we're hoping that we will get feedback, through that review process.

CHAIRMAN STETKAR: You know, I hope, through whatever channels, you can try to elicit that.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: Because a lot of the things that I'm hearing sound like it's a potentially quite complex process, that even though the demonstration may say -- show that it theoretically will work, in practice, there may be other impediments that you might not have thought about, and perhaps, some of the industry folks may or may not have thought about them, but it would be good to get their input, early.

MR. KURITZKY: I agree, and we're going to --

CHAIRMAN STETKAR: You know, beyond some -- you know, constructive input, beyond just saying, "Well, this is too complicated," obviously, we can't do this.

MR. KURITZKY: Right, so, I think there is

certainly going to be a certain kind of resistance. I mean, I, myself, being a PRA analyst, I'm somewhat resistant to it, and I'm thinking to myself, do I want to do this, when I'm doing my PRA for this one system?

You know, I've got a lot of resistance, myself, and I imagine that that would be a fairly common view out among the industry.

CHAIRMAN STETKAR: Well, but just resistance isn't necessarily helpful in this area.

MR. KURITZKY: I understand, but I'm saying --

CHAIRMAN STETKAR: People need to have some practical, you know, constructive input about an alternative approach.

MR. KURITZKY: Right, and that's why I think that because of that resistance, there is going to be -- we're not going to get as much feedback as we might otherwise.

If it was something a little more, you know, if it was something they would be a little more open to, but I still there by targeting EPRI, I think, we hope we will get some constructive input.

CHAIRMAN STETKAR: Okay, thanks.

DR. HECHT: Can I ask some questions about your Type I expert?

First of all, how will you know that if you take a Type I expert, let's just say, from the defense industry, who is both large scale command and control systems, that their knowledge about integration test practices are going to be relevant to your target system?

MR. KURITZKY: I think our attempt for Type I experts is to try to get people more associated with the nuclear field.

DR. HECHT: I see, okay.

MR. KURITZKY: I don't know how many there are going to be, that's the thing, it's the availability of such people.

DR. HECHT: Okay.

MR. KURITZKY: There is an identifying issue.

DR. HECHT: And I guess that relates to my second question, which is how will you assess the quality of that opinion?

In other words, if somebody says that this practice will decrease the defect density by 10 percent, but somebody else says it decreases it by zero percent, somebody says it decreases it by zero percent, somebody says it decreases it by 100 percent, what do you do?

CHAIRMAN STETKAR: There are -- and I'll point you to it, there are NUREGs on how to try to control for that, in an expert elicitation process, and make sure you pick up that. I don't have the NUREG --

MR. KURITZKY: It's the SSHAC --

CHAIRMAN STETKAR: It's the SSHAC methodology, whatever that number is.

MR. KURITZKY: Right, right.

CHAIRMAN STETKAR: It's not Bill Shack. It's seismic -- whatever.

MR. KURITZKY: Yes, it's SSHAC --

CHAIRMAN STETKAR: SSHAC methodology.

MR. KURITZKY: Yes.

CHAIRMAN STETKAR: It was applied for, you know, seismic analyses, but there is a fairly well developed and actually, bench-marked, you know, tested —- well, bench-mark, but fairly well developed method to try to control for that, to try to control the variability and given the variability, how to handle that within the process.

So, I'd certainly make sure that you reference that, because I believe that will probably form the basis for other initiatives in the agency, going forward, in terms of expert elicitation for a

variety of applications.

So, let's make sure you're tuned into that, and you know, you can't completely control that issue, but it does address that, that specific element of variability.

MR. YUE: This is actually the last slide. The last one is reference.

CHAIRMAN STETKAR: Cool. I'm sorry. Thank you very much. Anymore for Meng on the BBN, because now, I think we're going to come over to the statistical testing way of estimating input information.

Great, back to Louis, I guess.

MEMBER SIEBER: This one is a little more concrete, I think.

MR. KURITZKY: Is it possible for us to take like a two-minute break?

CHAIRMAN STETKAR: If you'd like to, yes, I was going to wait until about 2:30 p.m., but if it's appropriate now, let's do that, and we'll make it 15 minutes. Let's take a break now, and we'll come back at 2:25 p.m.

(Whereupon, the above-entitled matter went off the record at 2:08 p.m. and resumed at 2:25 p.m.)

CHAIRMAN STETKAR: Okay, let's reconvene

and see if we can have some semblance of schedule here.

MR. CHU: Okay, I'm presenting context-based statistical testing method.

As compared to BBN method, statistical testing were generated, in a sense, that you know, it doesn't have too much problem with the data, and certainly, the results is more measurable.

I will give a little background on statistical testing, black box testing or white box testing, and we propose to do -- use the context defined in the PRA, to do the statistical testing.

In particular, we tried to make use of PRA information to help us in developing the test strategies. Without doing the statistical testing, we can do some feasibility study by using a PRA, to estimate the number of -- total number of tests that's needed to demonstrate the software does not contribute too much to the overall --

MEMBER BLEY: Is this a good time to toss in what troubled me about that?

What troubled me is, it seems, the arguments about how that works, are arguments based on kind of traditional idea things, independent assumptions of identical and identically distributed

variables, and it strikes me that the kind of testing I read about, is really throwing what I'll call different conditions at the software, and seeing how it responds.

So, each case isn't identical. They're different, and using the kind of rules that say, I've done enough tests to know I'm good to 10 to the minus three or something, works if they're identically -- identical events.

But since they're not, I don't know that it really works the way you claim it does, because the cases that caused you trouble, might all be clumped over this in this part of the space, and you haven't even tested that part of the space.

So, it troubles me that the arguments on why this is adequate don't stand up to scrutiny.

MR. CHU: I think this is related to how you define your operational profile or test profile.

In general, we say, you know, the operational profile is the probability distribution over the possible input.

If we know the distribution exactly, and we take something from it, and feed it with the software, I think it's reasonable to claim that, you know, this allows us to estimate the software failure

-- failure probability.

But in reality, we may not know the operational profile well enough, and so, there is a difference between the test cases that you generate and the real data, doing systems operation.

Other than that, I tend to think -- but that's kind of an important assumption, that I think I point out in this morning's presentation, also, whether or not your profile use to generate tests, represent the true operating condition.

MEMBER BLEY: But yes, but you may have hit the essence of it, but it's not like, you know, the classic problems, where you have an urn with red balls and black balls.

If you take it that far, where you've got an urn with red balls and black balls and some green ones and some purple ones, and if the green ones and purple ones are the problem areas, and you never pick a green or a purple, you just get your black and red, your successes and failures of the normal set, doesn't work right, you know.

MR. KURITZKY: And that's the difficulty with software, because as you mentioned, with software, it's not identical testing.

In fact, the software, we don't want

identical testing, because it works for one exact input trajectory, we expect it to always work for that exact input trajectory.

MEMBER BLEY: Which isn't what you see with --

MR. KURITZKY: With random other stuff, right. So, in this case what you do is, you try to set up a test profile that draws as many balls as you can. You want to get them kind of spaced out.

As Louis mentioned, you want to try and get as close to operational profiling --

MEMBER BLEY: And right here, we're back to my trouble. If we don't know, what I've been calling failure modes, and we can call them something else, if we don't know what they are, then we don't know if we've picked the ones that we need to pick from.

MR. KURITZKY: Right, but what we -- again, it's more of a -- I don't want to say confidence game, that doesn't sound good. But it's a level of --

MEMBER BLEY: No, I think you're right.

MR. KURITZKY: But I don't want to say that, but it's a level of confidence, in that if I test -- if I understand how the software works, to the

extent that I do, and again, there is always going to be that delta, which is the stuff I don't understand, or I haven't identified yet.

But to the extent of my knowledge, if I do a structured, educated approach of setting up my testing profile, to try and capture as many different colored balls in that barrel that I'm drawing out, the more I can grab and with no failures, increases my confidence in the failure probability of that software.

MEMBER BLEY: Yes, it does, and if I made the software at the level I just heard you say, then I know about the things I'm calling failure modes, and that's my glitch.

So, I think we can go on. I just wanted to say, the argument, you're getting at the heart of the arguments now, and I didn't see those written down.

MR. KURITZKY: Okay.

MEMBER BLEY: The arguments I saw written down were, I'm going to go in and if I pick enough red balls, I know I'm okay, you know.

MR. KURITZKY: Right, no, that's not the point we were trying to make in any case, so, we'll make sure that we clarify that.

DR. HECHT: Can I ask a question, also, with regard to that same formula, which consisted of the three terms, which basically were the probability of that event occurring, which are -- or the challenge, I'll call it the challenge, the probability of the software working for that challenge, and then the probability of what I call the other factors, or the other failures, POF, as you called it.

Where do you consider the failures of the system itself, of the computer platform itself, and the probability of the software, I think you're talking about, if I were a TELEPREM system, you would be talking about the function block, but you're not talking about the platform software, and you're not talking about the underlying, what is sometimes called board support package, and that stuff that drips even into the semi-hardware area.

MR. CHU: We consider in the -- you know, the system may consistent on multiple processors, where we are considering the whole -- all the software on different processes, together, as a single system, and we're doing the test on that.

In that sense, the platform software operating system is part of it, too.

DR. HECHT: Part of what?

MR. CHU: Part of the system that we are testing.

DR. HECHT: Well, let's -- there may be some failure modes, getting back to the previous question, that are related to how long that software has been running.

For example, have you heard the term memory leak? Does that mean anything to you?

If the software is nearing just by virtue of its operating history, if it's been running for a sufficiently long time, it might be running out of memory because it didn't de-allocate all the memory that it took in.

So, that failure that occurs would be totally independent of the logic that would decide whether to do something.

Another example, earlier today, we were talking about close-loop control systems, not nearly safety trip systems, as part of the space of the systems that we're concerned about here, and those close-loop control systems have to operate over a long period of time.

So, whereas, I think in your method, you're just saying, do tests and do them one at a time, and look at the results of that test. I think

you also have to consider that there is an underlying failure rate, in terms of time, that also has to be factored in the model.

Maybe it just has to be another term in that model, but I don't think you can totally ignore just the time dependent platform failure rates, in a matter that you would consider, for example, simply hardware.

MEMBER SIEBER: I think another thing you can add to that is CPU overload.

DR. HECHT: Okay, CPU, the load might vary, even for the same exact function.

MR. CHU: Yes, we recognize that the memory issue is an issue. I guess we don't have very good way of addressing all the possible memory issues.

But if you look at the application software for reactive protection systems, probably the application software does not have too long of memory.

But there could be lower level software, for example, I think the patriot missile failure issue, where the accumulated error accumulate over time.

DR. HECHT: There are many such --

MR. CHU: That's at a lower level than -- I'm not sure, how testing can be done, to identify or

capture that kind of failure.

DR. HECHT: Well, perhaps, you can just take operating time and see the number of failures, just using the operational history of the platform, but I'm saying that the general issue that you're dealing with is software. It's not nearly logic entailed in the top level application that you're -- that you need to consider, and the failures of the digital system.

MR. CHU: Ideal, it says only, we want to test the whole thing, but the actual software at different levels and the actual hardware, I think later, we'll have -- you know, we'll talk about test configurations.

You know, what is the best test configuration, so that you can realistically represent system condition during its operation.

But there are practical limitations. You know, what equipment is available, that can be used in performing the test that we are planning to do.

DR. HECHT: Well, that is why you're -- you have a contract, because you're so smart, and you'll figure it out.

CHAIRMAN STETKAR: Let me ask a perhaps, more naive question.

The structure of this testing method, the information that we derived from the tests, or call it data, is the fundamental purpose of that information to provide data that I can use to inform the Bayesian belief network probability distributions, such that I can derive a failure rate, if I can use that term, for the software, or is the purpose of the exercise to determine a sufficient number of tests, such that I have confidence that the failure rate is no higher than 'x'?

MR. KURITZKY: That one.

CHAIRMAN STETKAR: Okay, why are we doing the second? For example, if I think of a diesel generator, in a nuclear power plant that has 57 diesel generators, the acceptable failure rate for a diesel generator would be much different than a nuclear power plant that has -- or one diesel generator.

So, information on what might be an acceptably small or large failure rate, doesn't really tell me anything about the real failure rate of a diesel, because if they're all identical diesels, each of those 57 diesels have the same failure rate as my one diesel.

So, why am I interested in developing information, just to give me confidence that the

failure rate, whatever that means, is not higher than 'x'? That's not, it strikes me, as solving the problem that we're actually trying to solve, for quantifying the reliability of software in the context of a risk assessment.

MR. KURITZKY: Well, what I would say to that --

CHAIRMAN STETKAR: Nor is it necessarily developing information that is consistent with what I thought was the prior distribution that was developing in my Bayesian belief network, which I thought was an uncertainty distribution, for what really is the failure rate.

MR. KURITZKY: Right, that actually -- that point is one that I don't know whether or not we've addressed that exactly.

When we talk about using the test result to bear out, or verify the prior, that is something that we're going to have to --

CHAIRMAN STETKAR: I didn't hear it say verify. I thought I heard, you were going to actually use it to update the prior.

MR. KURITZKY: Well, we are going to use it --

CHAIRMAN STETKAR: To give the update --

MR. KURITZKY: We are going to use it for the prior, but going back to the idea that we were going to use the statistical testing results as a type of benchmarking for how well our Bayesian belief network was able to estimate the prior distribution, that is something that would have -- I have to rethink that --

CHAIRMAN STETKAR: But you weren't going to do that.

MR. KURITZKY: Well, we were going to do that for the testing results. We were going to do that one case with the testing results.

But the one thing, we need more detailed benchmarking with other results.

But let's put that aside for the time being. What we're looking for -- you're right, in a pure sense of estimating the software failure probability, if we're just coming up and saying, "Okay, we need to do this amount of testing, to be sure that it's at least as good as 'x'," that doesn't get at the exact answer, as to what's the software failure probability.

So, again, it's a tradeoff. If we could come up and get -- do enough testing to give us confidence that this is the true failure probability

for the software, that's the best thing.

We can use that both in our PRA, and if someone who is doing a review, for acceptance of some submittal on a software can say, "Okay, as part of my determining the software quality, I know that the software reliability is 'x'," that helps them in their decision.

But the reality is, we don't -- in order -- the amount of testing -- the testing burden to try and do that, besides being impractical, may not even -- I don't know if it necessarily it's possible. You know, it depends on how many context and how much testing or how low of a reliability, or you know, how high a reliability the software has.

So, what we do is, using the risk information, we're trying to try and control the test burden, the testing burden, and so, this enables us to try and control the testing burden.

It doesn't give us the exact answer, to the software reliability question, but input into the PRA, if all the sequences that involve that software failure are all in the 10 to the minus seven range, then I don't -- using a software failure probability of 10 to the minus three, then I don't care, really, if it's 10 to the minus three, four, five or six.

CHAIRMAN STETKAR: Do you care if the total risk estimate for total core damage frequency is three times 10 to the minus eighth, and I'm going to use that PRA to justify extending the test interval on my protection system from once per year, to once every 30 years?

MR. KURITZKY: Well, that is where we use it -- that is why the process that Louis is going to describe is based on what the core damage frequency is for the plant, and what -- we don't identify exactly how you would determine what's an acceptable risk delta, okay.

We fill in Reg Guide 1.174 numbers, as kind of an example, but I mean, you'd have to determine for new reactors, certainly where we expect them to be lower core damage frequencies, that magic table that's in Reg Guide 1.174 may not be the table we want to use, that figure.

But the idea being, whatever you determine is to be an acceptable risk increase, if we can show that the software is at least reliable enough that it's not going to, you know, negatively impact or go over that acceptable limit, then at least we know that our pure rate, our estimates of core damage frequency for the PRA model are fairly robust for the value

we're putting in there.

CHAIRMAN STETKAR: But again, I'll come back to my diesel example, and my 57 diesel nuclear power plant, I might be able to do 15 tests to clearly demonstrate that the diesel generator reliability is better than, you know, nine failures and 10 demands, which is maybe all I need in a 57 diesel generator nuclear power plant.

But in my one diesel generator nuclear power plant, you know, it's a much different problem.

Now, perhaps your risk based context will solve that, but it's not clear to me that that's what I thought this research project was headed toward.

MR. KURITZKY: This proposal, I think initially, it might not have been headed towards that. I think with this proposal that's on the table right now, it is heading towards that.

CHAIRMAN STETKAR: Well, and that's the earlier concern this morning, are we getting ourselves into such a well defined little black box, because we needed to simplify this project, which to me, doesn't sound all that simple anymore, that we've lost track of what we're really trying to do?

MR. KURITZKY: Well, and I think --

CHAIRMAN STETKAR: And are there other

ways that we could accomplish that original goal? I mean, that is the --

MR. KURITZKY: Well, based on our review of the methods out there, we have not found a better method to accomplish that goal.

I think I agree with your statement that it's not as simple as you thought it to be, because I think even at this level, it's not that simple, and so, we're hoping that we can be able to do it, and your example of the diesel generators, putting aside the fact that if we had 57 diesels with the same software, then we would assume that software was going to fail them all, anyway.

But assuming there were 57 individual independent software, right, we're going to have to demonstrate for that case, that the software is just no worse than 10 to the minus one, or something.

Whereas, that other guy has the same software, the same diesel, but he will need to demonstrate that it's 10 to the minus five, or four or something on that order.

So, yes, it's specifically what this approach is, it puts it in the context of the risk of the plant, and someone who is relying on it more, that software more, for you know, or the system that uses

that software to keep this core damage low, is going to have to --

CHAIRMAN STETKAR: But it introduces a notion of an acceptable failure rate.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: Rather than an evaluation of what --

MR. KURITZKY: Exactly that is.

CHAIRMAN STETKAR: -- our best estimate of the failure rate with uncertainty is.

MR. KURITZKY: That's right, it does.

DR. GUARRO: But if you --

CHAIRMAN STETKAR: And that acceptable failure rate is an arbitrary notion. It's an externally applied constraint.

MR. KURITZKY: Yes, that's right.

DR. GUARRO: It can be that, or it can be something slightly different, though, because essentially, this is not, you know, it's not -- in concept, it's not different from when you have, you know, your generic failure rate of probability, from prior assessments versus what is the, we're going to call it plant specific information.

MR. KURITZKY: Right.

DR. GUARRO: And the question is, you

1 know, what do you —- you may end up believing more,
2 right?
3 When we are dealing with hardware, because
4 we have pretty strong confidence that one would say
5 diesel generators, they belong to a family of
6 equipment that we know pretty well, maybe -- I'm just
7 postulating that, okay.
8 You know, there is this faith that the
9 generic data may be almost as good as something that
10 you do specifically for your own diesel generator.
11 For software, that is certainly not true,
12 at least, you know, I don't think it's true, because
13 software is so multi-form that you cannot assume that
14 because something has been labeled, you know,
15 something is going to be behaving just like your
16 particular application.
17 So, the issue is now, what do you believe
18 more? You believe more, some parametric analysis or
19 BBN equivalent, you know, that has the quality, the
20 process, et cetera, et cetera, and comes up with a
21 number, based on expert opinion, or you're going to
22 believe more, your testing.
23 But the question -- so, when you do
24 testing, you know, obviously, the failure rate of
25 probability that you can infer from testing may not be

the true failure rate of probability. It's only what you're allowed to claim, based on the other testing that you do.

And so, the issue is, can you do enough testing to claim and number that is low enough to make you happy, or not, right?

So, I see it that way, you know, and because I don't believe in the generic data that much, I also think that that's probably what we're stuck with, you know.

CHAIRMAN STETKAR: Okay, yes, sir?

MEMBER BLEY: Early on in your soliloquy on -- which was brilliant, you said something I didn't like, and what you said was, we're going to see if our Bayesian belief networks can generate a prior that's good, or something like that.

And what I didn't like about that is, it sounds like this Bayesian belief network is some elaborate theoretical thing that generates things. It's really, can our experts generate a prior that's meaningful, and our experts include whoever builds that network, because that is a structure that you're trying to use, to find the best place to ask the experts the questions you think they can answer, to generate the final result, which you don't think they

can answer directly.

So, it's really kind of pretending that the mathematics is responsible for what comes out, but it's really the experts who are responsible.

MR. KURITZKY: Right, I wasn't trying to differentiate between that. I was just mentioning, there was that one bullet that we had in the previous presentation, where we said that -- when we had an early bullet that mentioned that it would be good go be able to bench-mark the BBNs, because that way, you know, we can validate the approach with some data.

But that was something that we said was out of scope and we don't know whether it could eventually be done, but that's something we, theoretically, you know, hopefully could be done at some point, elsewhere.

We said only in -- in our project, all we were going to do was, we were going to take the results from the statistical testing, and that would give us like one data point to compare against what the BBN generated, just to see, you know, how well it came out.

And an earlier point was, that was -- that's our plan, when Dr. Stetkar mentioned the idea about the fact that -- we're not actually coming up

with that failure probability, which then we could compare to what the BBN gave us, we're really just coming up with, well, it's at least as low as 'x', then that comparison is a little squishy then, you know, it's not quite —- doesn't give us quite what we were hoping to get.

CHAIRMAN STETKAR: Well, what you're coming up with is, if you run, you know, if I come back to my hardware analogy of diesels, if you run, you know, 100 diesel tests with zero failures, you can use that evidence for a Bayesian update, you know, you can certainly do that.

It will alter the shape of that prior distribution somewhat, and it's useful evidence. If you ran 10,000 tests with zero failures, you know, it would provide a stronger amount of evidence.

So, this whole notion of how many tests do I need to do, to gain confidence that the failure rate is lower than 'x', doesn't, from my perspective, mean an awful lot to me, you're just going to develop some number of tests, you know, and the number of tests that you can perform, it's going to be limited by the amount of resources that you have, and let's just leave it at that.

You know, why impose these arbitrary

notions of how many tests are acceptable for a certain failure rate?

MR. KURITZKY: Well, as far as the tests --

CHAIRMAN STETKAR: It's a possible statistics approach to life.

MR. KURITZKY: Well, I would go back to what Sergio had said, you know, we're going to have to -- we're going to do as many tests, to give us a number that makes us happy, that we can live with. I mean, that's — you know, we're never going to test an infinite amount.

So, the only thing about the number of tests would go to this idea of this risk increase. I don't know, if you're talking about risk-informed approach to the testing, where we were coming up with this delta of risk increase and saying, "Okay, if we're allowed this, that would help dictate to us, how many tests we need to do," that's the --

CHAIRMAN STETKAR: I'm not interested in, right now, a particular application for risk-informed upgrade to a digital I&C system, at a particular operating plant, under the auspices of Reg Guide 1.174.

I'm interested in that applicant who is

going to come in and say, "Look, I've developed a PRA model for my digital I&C system. I've quantified the unavailability for these particular functions, and I want to use this PRA to risk inform my tech specs," and this is a new plant design.

What do I know about our ability to derive numbers, and I'll call them numbers, that those folks may have used in their model, compared to, you know, the research information that we have available?

So, I need to think in that context, also, and there, I'm not looking, necessarily, at a risk increase over a current risk that I've quantified by whatever model, whatever other system is in there. They're saying, "This is, indeed, my core damage frequency with this information for my baseline PRA," and now, I want to extend the testing interval to, you know, infinite, and lo and behold, is a two and a half percent increase in risk, or something like that.

MR. KURITZKY: Yes, and --

CHAIRMAN STETKAR: So, that is what brings me back a little bit, from this notion of acceptable level, you know, the number of tests to develop an acceptable level.

I can understand that, within the context of a risk-informed upgrade to an existing system, for

which I have other bench-mark PRA results, that I'm then using as a comparison under, you know, 1.174 type analysis, and that's fine.

I mean, if that's sort of the way that this research project is morphing, I think we need to, you know, consider it within that context.

MR. KURITZKY: Right, and the only thing I would say is that if it was this other application you talked about, a new reactor coming in and trying to extend their tech specs, where, you know, just demonstrating, our approach just demonstrates that software is no worse than 'x'.

Theoretically, then, the licensee has come in with some kind of modeling approach to demonstrate their failure probability for their software.

Now, we would, again -- again, I go back to what Sergio said, and we're going to test -- you're going to test as much as you can reasonably test and you're going to come up with a number that's based on that amount of testing you did, and if that number is -- that we've done, shows -- makes us think that the number that the licensee is using is in that same ballpark, or not much better than the number we're okay with, then we can say, "All right, well, that seems reasonable."

If their number comes up to be much lower than the number that we were able to generate with our reasonable amount of testing, then it's a question, and then we have to start looking in more detail, because you have to say, "Is that just because we didn't test enough," or was there something about the different approaches that is the cause for the difference in the answer?

MEMBER SIEBER: One of the difficult concepts in all of this, to me, though, is that if you look at probability of failure for mechanical equipment, as an example, it has a life time and you turn some pump on and it runs and runs and runs, until it wears out, tolerances increase.

And to me, that's a pretty predictable, pretty reliable thing that you could assign a probability of failure to, put it in a risk assessment.

The instrumentation, though, you've got a fault that's in there from the beginning, and the question is, when do I get to a sequence that hits that fault and puts the plant in jeopardy, and I don't think that has any fate or relationship to time, what so ever.

CHAIRMAN STETKAR: I think that comes back

to Dennis's question about, you know, the sampling across what you characterize as the operational profile --

MEMBER SIEBER: Right.

CHAIRMAN STETKAR: -- is, are the tests that you're performing and the statistical sampling of those tests adequate to capture --

MEMBER SIEBER: That's exactly right.

CHAIRMAN STETKAR: -- all of the conditions that you might anticipate across that operational profile, and perhaps, not knowing what those conditions are.

MR. KURITZKY: Right.

MEMBER SIEBER: Yes, and the --

MR. KURITZKY: And the uncertainty --

CHAIRMAN STETKAR: Not only yellow balls and red balls and black balls and green balls and purple balls, but you know, there is a cube of sugar in there, you didn't realize it was even something that wasn't a ball.

MR. KURITZKY: Right.

MEMBER SIEBER: And you have to go even a step further, if you perform the tests one at a time, you aren't really testing the system.

It's when you have an event that causes a

lot of strain, lot of computations on the computer systems, that causes an overload, causes mis-directions and what have you, off-normal conditions, and I don't see anyplace in here, where you actually are testing for that.

MR. KURITZKY: Actually, that's actually what we are trying to test for, specifically, with this approach, and it goes back, I think to stuff that was in -- ASCAs work, where their context-based software reliability method, where they've shown that typically, the bulk of the problems occur with off-nominal conditions.

The nominal conditions, the designers are aware of this stuff. The software guys are programming, and they're very comfortable with that, it's these off-nominal conditions that they didn't think about and that's where they run into trouble, and specifically, that's what the CSRM approach tries to do.

It tries to focus on this off-nominal conditions, the saving grace being that the off-nominal conditions have a lower failure -- you know, have a low occurrence likelihood, and so, therefore, you don't have to demonstrate -- hopefully, so, you don't have to demonstrate as high a reliability --

MEMBER SIEBER: That means if you are in trouble, you don't get in trouble very often, so, you don't have to worry that the trouble you get into is fatal, okay.

MR. KURITZKY: Well, the point is that from the risk point of view --

MEMBER SIEBER: Anything can fall apart.

MR. KURITZKY: -- your conditional failure probability for a low, doesn't have to be as low for a low likelihood occurrence, and that's the same concept that we're trying to use, also, with the risk based approach, and looking at the -- eventually, to identify the boundary conditions for the test cases, we're trying to consider those off-nominal cases, where you do have all of the things going kaflooey at once, and the system has to deal with that, and that's exactly what we're trying to model here.

As far as the time reaction, this doesn't have that time factor. This does do -- because it's not a failure rate, it's a demand failure, so, it takes these cases and it looks at the different conditions, and again, it's sampling the balls from the barrel type of thing, and it doesn't fail with time, it's just how well you cover that whole operational profile.

And the limit, you would cover the entire operational profile exactly, in which case, then you would have perfect software, because everything you found, you would have fixed and there would be nothing that could go wrong.

In reality, we can't -- no one tests the whole operational profile. So, there is some probability that there is something that could go wrong, but you haven't identified and haven't fixed, and so, all we can do, we don't get, on like a random failure probability, where we have some statistical methods to give us some likelihood of failure, based on the various random failure mechanisms, we have -- you know, how confident are we that we've drawn enough samples of balls, knowing that we're looking at the cases where we think there would be the most trouble, that we -- you know, we haven't run into any problems, so the more we do that, the more our confidence goes up.

And then at some point, we say, "This is what our confidence is," and that's the value we're going to use for right now.

DR. HECHT: I would just like to re-emphasize the fact that all of that is based on the assumption that your platform is there and working.

I still think that in addition to that testing of the system response to individual challenges, you have to consider the unavailability of the system.

MR. KURITZKY: When you say 'system', you're --

DR. HECHT: As part of it.

MR. KURITZKY: -- talking about the --

DR. HECHT: I'm talking about the operating system, the device drivers, the libraries that you're using to create your function blocks, all of which are running all of the time and all of which don't know whether you're running a test or running under normal conditions.

MR. KURITZKY: Right.

DR. HECHT: But which fail unexpectedly and dare I say, randomly, and have a certain failure rate that -- and therefore, have a failure probability that can take your system down when you need it.

MR. KURITZKY: And I think that some of that, we hope we will capture in doing the integrated testing, as Louis was talking about, having the right hardware, the right test configuration with hardware and software.

Some of it, as you mentioned before, some

of it, we may not just be able to capture, and just going to have to acknowledge that that's the limitation.

DR. HECHT: No, you just assign a probability, just like you assign a probability of the hardware failure.

MR. KURITZKY: Right, or it's in our distribution that may have to be -- you know, a limit on the upper bound of the failure probability, too. So, well, anyway.

CHAIRMAN STETKAR: Let's let Louis try to get to slide three.

MR. CHU: Okay, we're going to try to apply this method to the examples, as later, I will discuss a little bit about the examples, in terms of testing.

A little bit of background, the National Research Council had recommended that random testing and expert judgment be used in assessing software failure probability, and they say, including software failure in a PRA of a nuclear power plant is preferable, though it is our statement supporting quantification of software failure, using testing.

In 2004, ACRS Subcommittee emphasized the importance of system-centric view of software failure

and importance of considering context in which software failure can occur, and I have already mentioned Sergio's work for NASA is context-based software risk model to CSRM.

There, he used the PRA model to define the context for performing statistical testing.

A little bit software testing performed for licensing purpose, for licensing purpose, different types of tests are performed. For example, unit test, integration tests, system tests, factory acceptance tests and site acceptance tests, these tests objectives is to show that the bugs can be -- again, identify some bugs, so that they can be fixed.

And to demonstrate that the software can perform its intended function, some test requirement may call for coverage, such as every statement needs to be exercised.

A key point I want to make is that these test results, in general, cannot be used to quantify software reliability, because they are typically designed for specific purpose, and the test cases are -- were not sampled from the operational profile of the software.

Therefore, looking at all these tests performed for licensing purpose, we feel that

statistical testing, that we propose to do, can be some tests that can be done as a part of the factory acceptance test, or after the factory acceptance test is done.

A little bit of literature reviewed, there are theoretical research done by Miller, May, Littlewood and Kuball, mostly talking about, you know, the theory on how you do the test and how you quantify the reliability.

While May's work may -- John May's work related to Sizewell B, I think they have actually performed some tests, statistical tests on Sizewell B systems, but his paper itself doesn't really give us too much information about the statistical testing done on Sizewell.

In addition, I think the Kandu reactors, they have some kind of requirement to do tests, to demonstrate some kind of goal like 10 to the minus three failure probability.

There are some research done on white box methods, that were done by Zhang, it's an MIT thesis and Yang, who is NRC staff and the white box method, he need to keep track of the paths and nodes in the software, in doing the tests.

Costs associated with statistical testing

may be significant, thinking of the test equipment needed, resources needed to carry out the testing, and the number of tests that need to be performed.

Black box testing, basically, as I described earlier, you define an operational profile from which you sample the test cases, and you feed the test cases to the software and then determine the software responded correctly, by generating the signal, and if we use a uniform prior distribution, assuming the failure probability is beta-distributed, then this expression gives you the mean failure probability given this 'n' tests with no failures.

So, black box testing methods seem to be the most straightforward method, in terms of testing.

White box testing method, you need to keep track of the coverage, how frequently a path or a node in the software is visited, therefore, it requires a lot more effort in performing the test, and in a white box method, basically, you estimate how frequently the node or path is visited and then, you test the particular path or node, and determine a failure probability, and you calculate the average failure probability for the system.

MEMBER BLEY: I'm just curious, because I've never been involved in this stuff, but this white

box testing idea, it seems to me, to be able to do that, you have to build code inside the code, to track all these executions as you jump around through all of these nodes.

Is there any history of the additional code you have to write, to run the test, actually interfering with the test? Is that a big problem?

MR. CHU: I think keeping track of that kind of information probably required that you enter some information --

MEMBER BLEY: I would think so.

MR. CHU: -- showing the —-

DR. HECHT: It's called code instrumentation.

MEMBER BLEY: That makes sense.

DR. HECHT: Yes.

MEMBER BLEY: But has there been experience with that actually creating problems in an of itself, so that the test results are affected by the test instrumentation?

DR. HECHT: I can tell you what's done in the Avionics industry, and that is that when you want to do this structural testing, you want to test every statement, for example, every condition.

It's very difficult to create test cases

to do that, so, you just create a large number of sample test cases, run them through, the code instrumentation tells you how much of the code has been covered.

You take the instrumentation out. You run those same tests, you get a notion of success, presumably, they all pass, and then for the residual paths, which you weren't able to cover, then you kind of do something to create additional test cases to cover what you did --

MEMBER BLEY: Drive it through.

DR. HECHT: -- and you may do this several times.

MEMBER BLEY: So, you actually get a check on yourself, as to whether you've had an impact or not.

DR. GUARRO: You see, one of the objections to try to model, you know, software structures is, you know, the combinatorial explosion thing.

MEMBER BLEY: Sure.

DR. GUARRO: Well, here, you have the same issue, right. If you were to test all conditions in all the possible branches of the conditions, you're exactly in the same combinatorial explosion that you

had with modeling the system.

So, conceptually, I think you have the same dilemma to face.

MR. CHU: Yes, that's exactly the difficulty with the white box testing method. It requires additional efforts, a lot more.

DR. GUARRO: Yes, the flip side of that is black box, you know, that is something, and that applies to everything.

So, it's the ostrich solution to the problem, you just stick your head in the sand, and then you could declare a victory.

DR. HECHT: Yes, that's where you get into the sugar cubes with the black balls and white balls.

MR. CHU: In case of statistical testing or black box testing, the operational profile is very important part of the testing, because it defines the possible input case that can be generated and used in the testing, and how you can define the operational profile and characterize it is an important issue.

In general, you know, you may not know certain input can trigger a failure and somehow, you missed that when you defined your operational profile, then you're going to miss it when you do the test.

MR. KURITZKY: To put it in Dr. Bley's

words, it would be like, did you cover the case, where you reach in, and in real operation conditions, would you reach in and grab the sugar cube?

It may be that the sugar cube exists in there, but the real operational profile is that you would never grab that sugar cube, and you're okay, but you don't know that all, because you're never going to completely test the operational profile.

MEMBER BLEY: Well, and sometimes, data come in that's not part of the operational profile because of failures in sensors and other things, that generate data that you wouldn't expect to ever see in the real world, but it gets generated.

MR. KURITZKY: Right, but that actually should be covered in your -- I mean, that's considering the off-nominal conditions, you really should consider that.

MEMBER BLEY: Okay, you should be able to pick that out.

MR. KURITZKY: Right.

MEMBER BLEY: Okay.

MR. CHU: So, in general, it's hard to estimate what you're missing, in terms of, you know, trying to come up with the operational profile for testing.

I think in Sergio's work, he used expert opinion to come up with some kind of estimate, some kind of increase in probabilities, due to the difference between the actual operating profile and the test profile.

But it's based on expert judgment, you know, how reliable are the numbers, is somewhat questionable.

Similar research provide guidance on how to randomize when you generate test cases. For example, Littlewood and May, they each have some publication that would suggest how statistical testing can be done, and May introduced the term 'trajectory' in the space of physical parameters.

Basically, a trajectory is the path that a particular scenario occurs in the space of physical parameters, and you need to use thermal hydraulic model, to generate such a trajectory. The main reason that you need to do that is read sensors from the plant, and the sensor readings are physically inter-related by the physical processes.

So, by modeling the thermal hydraulic part of the plant, you can automatically account for the dependency among the different sensory inputs.

CHAIRMAN STETKAR: I guess I'd like to

understand, you know, you've introduced the notion, both this year and last year, of linking now, a simulation of the plant software to the simulation of the plant thermal hydraulic response, which is starting to sound to me, an awful lot like a full scope simulator of the plant.

How does a thermal hydraulic model of the plant capture this notion of off-normal conditions in terms of the plant response?

MR. KURITZKY: Well, I would --

CHAIRMAN STETKAR: That might be introduced, for example, not from the plant response to a turbine trip, but from plant response to a fire, the cause spurious signals for input sensors, it could be outside of the ranges that any thermal hydraulic model might ever predict, because the thermal hydraulic model know that those parameters would never get that big.

You can't ever have pressurizer level, you know, at 100 percent, for any of these scenarios that you've thought about, and yet, if pressurizer level was 100 percent, the software might respond differently, than any of these thermal hydraulic models could force it to.

MR. KURITZKY: Well, again, I think that

the boundary conditions you get from the event trees in this scenario is --- you know, feed into your thermal hydraulic calculation, and that's the identification.

Now, you make an interesting case with the fire type scenarios, because then you're getting really off --

CHAIRMAN STETKAR: Well, but it's a problem that we do face, when we start looking at fire analyses, because very often, we find that the logical structure of the PRA model, for good thermal hydraulic analyses, forget anything else, that the logical structure the PRA model has not anticipated the types of failures that we can induce, for example, and I use fire, but it could be a seismic event, it could be any number of other conditions, like I'll use fire, only because it's an example I'm familiar with, spurious actuation type things.

Very often, the -- you actually have to go in and either change the structure of your logic models, or appropriately trick the logic model for it to respond to the combination of input conditions that you're creating.

So, using a pre-existing logic model to inform how you -- you know, as a basis for risk

information, about the scenarios that you're going to then challenge your software testing scheme, may still not capture those off-normal conditions, because you're still constrained by somebody's presumption about what the range of failure conditions are that PRA model will handle.

And if it's PRA model for a previously -- you know, a currently operating plant that has only an analog system that was developed by somebody, you know, 20 years ago, there are even further constraints about how useful that model might be, you know, from this perspective.

I'm curious, if you've thought about that.

MR. KURITZKY: I don't know whether Louis has thought about it. I haven't gone -- I certainly haven't thought to that level of detail.

I think that as with anything in the PRA, you're only going to include those things that you've thought about, and there is always going to be things you haven't thought about, and are not well captured in your model, and this is a particular case where I think for the basic internal events and more what's called normal events that we -- the idea of using the boundary conditions from the PRA to generate the -- or using the PRA to generate the boundary conditions for

the thermal hydraulic test cases seems reasonable.

There is a lot of other concerns that we're going to get into, but I mean, it seems reasonable in that regard.

But like you said, there are going to be cases, whether it be seismic or whatever, that are going to even be a little more complex than what we were thinking about, just from the basic turbine trip, et cetera.

CHAIRMAN STETKAR: Well, but I use those only as examples, because I keep coming back to this notion about understanding what that operating -- the first bullet on this slide, the operational profile needs to be defined, such that random samples can be taken from it.

And if the operational profile, as Jack has pointed out, can sometimes include off-normal conditions that the PRA modelers, who developed that particular logic structure 'x' number of years ago, haven't anticipated, then the PRA models themselves may not capture the full range of that operational profile, and it might not capture it in ranges where you have no understanding of how the software will actually respond.

So, in that sense, we're not sampling from

-- I think of it as tails of the distribution, in that sense, we're not sampling from those tails of the distribution, because we don't even know -- have the ability to define where those tails are, from the construct of the model that we're using.

MR. KURITZKY: Right, and I think that --

CHAIRMAN STETKAR: So, that again, you know, I don't want to -- I use the fire as an example of, you know, how we've had to trick models in the past, but it's that sort of notion.

MR. KURITZKY: Right, and I think this goes to the limitation. There is going to be limitations to what you can do.

It goes right back to what Dr. Bley mentioned, about getting the purple balls or the sugar cube, or whatever. That's all part of that operational profile, and there is two aspects.

You're only going to be able to -- the operational profile you're going to try and generate from is only based on your knowledge of how many things you can conceive of that can go wrong.

Like, right now, we can say, "As we look to do our operational profile, besides essential events, maybe want to throw in some external events," we have to kind of envision how the impact of those

external events could affect the various parameters of the plant, you know, to what levels.

There is obviously going to be some limit. I mean, there is a finite number of values a certain parameter can have. You may peg the scale or whatever, but it's finite.

It's all going to be a question of tradeoffs and limitations. We're never going to be able to identify the exact entire operational profile and test for it. We're going to try and do as complete a job as possible, trying to take into consideration, as many things as we can envision.

But there is -- it's not going to be a very comfortable answer, but it's going to be the fact that the more we can test, for the more range of things we can think about, the more confidence we have in the number, but there is going to uncertainty.

CHAIRMAN STETKAR: The only reason I bring it up, Alan, is that I see this project progressing into a very elaborate simulation model.

You know, we're not talking about linking a thermal hydraulics model to, I guess, real software and running cases of that integrated model --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- using another PRA

model to try to inform that stuff. That starts to get pretty complex, pretty resource intensive, and in the same notion of more detail is always better, is that true?

Is there a different way to capture the range of an operational profile thinking different about how the world works, if can be that broad, possible, rather than marching down this process where more and more complex simulation is the solution to the problem?

MR. KURITZKY: I have that exact question.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: But I don't have the answer.

CHAIRMAN STETKAR: Okay, as long as the research -- I don't see that question being asked as --

MR. KURITZKY: Well, it's asked. You don't see it, because it's not asked in this particular -- in the documents, because we're asking that, as we try to determine how to proceed.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: I always, in my mind, being a PRA guy, I always want a simple solution.

As I mentioned before, I don't want a $5

million study to answer the question for one software package in my nuclear plant. You know, I want something that is good enough to get me by, that I can include this stuff in my PRA and I can make regulatory decisions with it, and to the extent I can use traditional fault modeling and something much more practical and reasonable to apply, that's what I want.

Right now, we haven't put our finger on anything like that. This is what we've identified as to how we think can address the problem.

I think if someone -- as we're going through this process, if someone can bring up something that says, "Hey, that's a much simpler solution that might work for us," I'm all for looking into it, full steam.

But right now, this approach that we've thought about, that looks like it could do the job, even that is questionable, but we want to go through and see how well it can do the job.

There is two things. One, can it even to the job, because if it can do the job, can we then take what we've learned from that and somehow, makes things more simple, more cookbookish or whatever, for people going forward, so they don't have to go do this whole thing again.

I don't know, but I don't think we're ever going to really know, until we start to actually move through it.

But I have no problem saying, "Hey, we've spent 'x' dollars and 'x' effort to go down this path. It's looking hairier and hairier," now, all of the sudden, we've just heard about this other approach that might be a lot more reasonable, let's take a look at that, too, whether it's --

CHAIRMAN STETKAR: Well, that's my whole point, is, where do you get to that point, and by the time you get to 'x', whatever the 'x' dollars is --

MR. KURITZKY: Right, but it's --

CHAIRMAN STETKAR: But I mean, if you honesty feel that you've exhausted sort of, the current state of knowledge on the possible paths, then that's fine. You know, it's -- we've not been privy to all of the details of the work you've done to examine, you know, alternative ways of either defining the operational profile or, you know, estimating the failure rates.

MR. KURITZKY: Right, and so --

CHAIRMAN STETKAR: So, you know, that's a degree of confidence.

MR. KURITZKY: Right, and the big thing

is, that alternate path, I mean, if we don't have an alternate path to work off of, then I believe even if it's expensive or resource intensive, then we should go through and at least, if we can, do at least one way.

So, at least we say, "Here is at least one way it can be done," but if we could find some alternate, more practical path, like I said, I'm all for that.

The only way -- we haven't identified something and we haven't found that solution, but that was the whole purpose, why this project, this report in particular and previous ones, we always put it on for public comment, for targeted external peer review. We tried to get it out there because, we're looking for something --

CHAIRMAN STETKAR: Yes, you're hoping that somebody comes out of the --

MR. KURITZKY: Right, exactly.

CHAIRMAN STETKAR: -- woodwork and says --

MR. KURITZKY: And says, "Hey, you didn't think about this."

CHAIRMAN STETKAR: -- "Here is this approach."

MR. KURITZKY: Right, exactly, so, I'm all

in favor of that.

CHAIRMAN STETKAR: All right, okay.

MR. KURITZKY: We just haven't found that yet, so, this is --

CHAIRMAN STETKAR: No, and that's fair. I mean, you know, if there is something out there, as long as you're giving people ample opportunity to offer that up, that's about all you can ask.

MR. KURITZKY: Right, and that's why we always --

DR. GUARRO: I would say the one piece of the puzzle that is missing, in general, is the -- some classification of the types of software functionality that you're talking about, because I think that the issue of whether you can use a simpler approach, is very strongly related to how complex your software is.

Because you know, in other words, you could have a cookbook approach that says, "Well, if I do these things, I feel pretty good that I've covered the possibilities."

Well, that is if the design of your software has been constrained to be simple, to a certain degree. You know, I cannot give a definition of simple, but you know, if it is -- on the other hand, if it's more open-ended or different degrees of

complexity, then it's more and more difficult to have a cookbook that says, "Okay, you do these three things and you're covered."

So, it's -- you cannot -- the difficulty is not -- the difficulty is to be able to give an answer that is general, without at the same time, having to put constrains on the design process, itself.

CHAIRMAN STETKAR: Okay.

MR. CHU: All right, about operational profile, I just want to explain a little.

Since we are doing context based testing, therefore, for each context, we have to define an operational profile.

So, when you talk about -- you have the fire or whatever scenario, it's just defined in other context for doing the extensive testing, because you tell me what's the effect of the fire. I tried to do this software testing, based on that context.

CHAIRMAN STETKAR: Okay, go on.

MR. CHU: Okay, next slide, white box based statistical testing.

It takes into consideration of internal structure and tries to make sure different elements in the software is covered.

Certainly, white box method will give you more detailed information about individual parts or nodes, paths, but additional resources are needed to keep track of that information, and I will say when you test -- do the testing using white box method, you may have to institute additional -- you may have to, in fact, change the original software, then.

You know, then your test is not as realistic as the test without changing the original software. So, that can become another limitation of the white box method.

In May's work, he derived an expression for the means failure probability, in terms of the number of tests and partitioning factor. That is, he assumed that the input space is partitioned into 'n' partitions, and you do tests for each partition, and this expression gives you, how you can calculate the mean failure probability.

By looking at this expression, you can see that. If you want to show that the overall system has certain probability, you need to -- say, using black box testing, you need to do this number of tests, and by looking at this expression, you will say for each partition, you have to do the same number of tests, in order to arrive at the same means failure probability.

So, in that sense, when you do white box testing, you may end up needing about -- more tests to be done, even though it gives you more detailed information. That's in May's work.

I think similar observation can be made about white box method that expressly accounts for path or nodes.

In Zhang's work, the MIT thesis, she found that in her study, she was not able to consider all the possible paths of the system that she analyzed.

So, these are kind of indications of some kind of difficulty with white box method. It required more resources and it's maybe more difficult to do, even though it gives you more detailed information, and later, I will explain more about what we propose to do, what we call the risk-informed test strategy.

Effectively, we are looking at doing tests for individual sequences, or maybe you can even think of, they have more detailed information, that is allowable.

So, this kind of thinking is somewhat similar to white box testing, in the sense that the number of tests needed to implement the test strategy will be high.

The only thing that can possibly help in

reducing the number of tests needed is considering the risk information, and basically, you can somehow demonstrate that the software failure probability is at -- estimates the upper bound of software failure probability, such that the overall contribution of the software to the risk is low enough.

The importance of test configuration, test configuration needs to be as realistic as possible. For example, if you do testing of the software, but you don't run the software on the actual hardware, then you may miss some interaction between the hardware and software.

Therefore, one can think of different ways of testing the software, and using different test configurations, they represent different degrees of realism and in general, you know, the most realistic test configuration gives you the better results, but in practice, there are practical limitations, that is, in case of when we are planning to do -- we probably have to use one of the test configurations that we can come with.

But in general, we can say we demonstrate the approach, and then the plan, or the vendor that has better equipment, they may have the test configuration more readily available, they can

possibly do the test more easily.

An example of importance of test configuration can be explained in terms of a Oconee upgrade.

Originally, I think Oconee tried to use the SIVAT, which is a tool for testing the software, without the actual hardware, and I think it's NRC's opinion that they don't consider that kind of testing as an acceptable validation test tool.

So, they end up using the ERBUS test machine, actually test the both software and hardware.

Of course, it's not necessarily easy for the NRC staff to do these tests, unless NRC want to, say, build some kind of test center, generic to testing software, but I don't know that NRC has such plans.

Using PRA to define context for statistical testing, it's pretty -- I guess, the thought is pretty straightforward. PRA certainly identifies places where actual systems or protection systems are needed, and it could be the reactive protection or it could be the ESFAS system.

PRA event trees certainly specify where the systems are needed, and therefore, we call that -- I guess we call those sequence as the context for the

software, for testing the software.

MEMBER BLEY: As I read the report, I was a little concerned, here, that if you just use the context that's modeled in the PRA, to define the conditions under which to test the software, there might be context conditions under which the software might fail, that aren't well defined within what we model in the PRA.

But those conditions could still come up, when you're responding to the things that are modeled in the PRA.

So, did I misread, or are you limiting the context for testing to just those things defined by the PRA model?

MR. KURITZKY: I'll take the first stab, and then Louis can fill in any gaps that I miss.

But the idea is that the first level of defining the context is from the PRA, which would define what things have failed, or you know what other events have occurred in the path.

Then within that context, defined by those boundary conditions, the idea is, we're going to -- you're going to have distributions of things that you're going to vary. That's where you're going to get all the different -- the operational profile that

Louis mentioned for each context. That's where you're going to vary, for instance --

MEMBER BLEY: And that won't be strictly limited to things that are identified in the PRA?

MR. KURITZKY: No, it will --

MEMBER BLEY: Okay.

MR. KURITZKY: All the different values will be -

MEMBER BLEY: That's what I was hoping you would say.

MR. KURITZKY: -- for all those different hydraulic parameters are all different locations and sizes of LOCAs, if you say small LOCAs --

MEMBER BLEY: It didn't sound that way, when I read the report. It could be because I read it quickly, but you might take a look back at Chapter 6, in that area.

MR. KURITZKY: Okay.

MR. CHU: We can consider that all as part of the operational profile. You know, what else can be varied?

MEMBER BLEY: If the report says it, the way you're saying it now, I think that's good, but double-check Chapter 6 and see that it does --

MR. KURITZKY: Even if it does say it,

we'll beef it up.

MEMBER BLEY: Yes, okay.

DR. HECHT: I have a related questions. Perhaps, the opposite question, and that is, how many contexts have you identified from the PRA? Do you know what that is, particularly, in light of the difficulties you pointed out in setting up the tests and the expenses of running the tests?

MR. CHU: We haven't tried to identify, but typical PRA probably, you know, has maybe few hundred sequences, and probably every sequence, it needs some protection function. In that sense, there could be hundreds of sequences.

Of course, we're not going to test every sequence. We will just do a demonstration, showing, you know, select, say, a couple of sequences and show how statistical testing can be done.

But without doing tests on the hundreds of sequences, we propose to estimate, what is the number of tests needed? What is the total number of tests needed --

MEMBER BLEY: Even though you're not doing them?

MR. CHU: Right.

MR. KURITZKY: Let me step up, one second.

MR. CHU: That's what I called it --

DR. GUARRO: I suggest something, which is, you know, you shouldn't necessarily look at the hundreds of sequences.

I think you should look at what "end states" for the software are relevant.

In other words, you should treat the PRA sequence -- you know, PRA models as a way of identifying a set of context, and that doesn't necessarily -- it's not a one-to-one mapping with the sequence. Hopefully, it's a many to few mapping.

MR. KURITZKY: Right, that was the point I was just going to make, is that in fact, we're going to go through all the sequences, and we're going to do a feasibility study, where we look at a PRA to see how many sequences theoretically, there could be, how many contexts.

But the reality is, for this to be at all practical, there is going to have to be substantial binning. We're going to have to be able to condense the sequences into certain bins, and say, "Okay, these all act similarly."

CHAIRMAN STETKAR: And I think I'd caution that the term 'sequence' doesn't necessarily mean anything. I don't know what the definition of a

sequence is, because depending on the pictorial structure of a particular PRA model, you might have a couple of hundred, I'll use that very distinct term, sequences as individual paths through what I might define as my event trees.

Somebody else might define -- develop a PRA for the same plant, with a different logic structure that might have a couple of million paths through event trees, simply because of the way the event trees are structured.

Both of those models might have supporting numbers of cutsets, which indeed, are the true logical representation of the combinations of failures, which you can equally characterize as sequences, and those might number in, you know, the billions.

So, be careful, what you're characterizing as a sequence, because in many cases, that's an artifice of the particular pictorial representation of the logic structure and where a particular analyst has decided to define the interface between an event tree logic versus a fault tree logic.

MR. KURITZKY: Right, and actually --

CHAIRMAN STETKAR: So, you know, just be careful that way, because this binning issue does become very, very important, when you drop down to

that fundamental cutset level, if you will.

MR. KURITZKY: Right, and now, you've put your finger right on the issue that keeps me up at night, is the fact that really, as you just mentioned, it's at the cutset level where you're really defining --

CHAIRMAN STETKAR: Because everything else is an artifice of --

MR. KURITZKY: How you do the job --

CHAIRMAN STETKAR: -- my representation of an event tree versus a fault tree, versus yours, versus, you know, any number of analysts for the same plant.

You might ultimately have precisely the same cutsets and yet, wildly different estimates of what you'd define as a sequence.

MR. KURITZKY: Right, and actually, I think the report identifies, I know at least as we were writing it, it was -- and I don't know if it's still in there, I think it is, that we talk about sequences, but it's really cutsets, where you might have to go --

CHAIRMAN STETKAR: Yes, I didn't see that. That's -- I might have -- I missed it, also.

MR. KURITZKY: Right, but the point is,

once you get down to the cutsets, you better be able to do one heck of a lot of binning, because you are quickly way out of the realm of practicality -

CHAIRMAN STETKAR: Right.

MR. KURITZKY: -- if you can't do substantial binning.

CHAIRMAN STETKAR: Right.

MR. KURITZKY: And that's why that's my biggest concern --

CHAIRMAN STETKAR: And that's a real challenge, that's right.

MR. KURITZKY: Yes, that's my biggest concern, with the entire approach.

DR. GUARRO: There is a flip side, which is, of course, if you use the traditional PRA, to determine the context, then that's good, but there is one element that is missing there, or at least, is not explicitly -- this information is not explicitly contained, which is the timing of all events.

The timing of the pieces of the cutset, which for software testing, is important.

CHAIRMAN STETKAR: And that is what I thought I heard Alan say, they were going to try to capture, I think, maybe I misheard you, in terms of the range of -- for a particular -- if I can define a

context as a grouped set of cutsets, for the moment.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: I thought I heard you say that within that context, you are then going to look at a distribution of what might be the operational profile --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- for that context, which I would presume would capture variabilities in the timing, which you're absolutely right, Sergio --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- the static PRA model doesn't --

DR. GUARRO: Doesn't contain --

CHAIRMAN STETKAR: -- contain that information, explicitly.

You can infer ranges, given that context of your group, but --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- but it doesn't have that timing information.

MR. KURITZKY: Right, and that goes right with what Dr. Bley said before, that we want to -- there is a lot more than what's in the detailed PRA, that we're going to be varying.

The point is, though, that as we -- because see, there are a lot of such things. Besides the fact that we just talked about how many god-awful number of context you might be looking at, then within each context, the things we have to vary for that operational profile can quickly grow on you, too.

Take an example of a small-break LOCA, where you want to vary the size, the location. If you want to vary the timing of the various things coming out or coming off, or what states the plant is in when it occurs, I mean, there is -- you can certainly envision all kinds of different variables that you'd want to range over, for your operational profile, and it quickly becomes, you know, an intractable problem.

So, again, we're going to step into this, we're going to wade into this murky pool and we're going to try and see whether we can constrain it enough that we can be practical in getting something done, but I'm not promising anything right now.

CHAIRMAN STETKAR: We got that.

MR. CHU: Okay, I think we discussed a lot about context. Let me go onto the next one, risk-informed test strategy.

Basically, we tried to use PRA information to help us to develop a strategy for testing. There

are two basic ideas.

One is other failures in the PRA sequence or cutsets, may reduce the number of tests needed. To demonstrate that, the sequence of cutset, frequency is thrown off, or to -- the needed software failure probability is -- though it may not require very large number of tests.

And this probably is especially true. Say, you take reactor protection system. For those plants who sequences are not important contributors, then maybe it's easier to show that, you know, you don't need to have too many tests, to show that software does not contribute too much to the plant risk.

MR. KURITZKY: But let me make a point. This is the point where we diverge from calculating software, it's like estimating software failure probability, to just demonstrating that it's low enough, and at this stage, we could theoretically, if the sole goal was just to -- we want to know what -- estimate the software failure probability to the best that we can, we would just go ahead and test.

We would use the PRA to still come up with all the context. Each one would be tested up to some limit of what resources allow, what time allows,

whatever is practical, and we would come up with our best estimate of software failure probability for each of those context.

Now, at that point, we can say, "All right, well, if that -- that could still be an onerous amount of testing that has to be done," and then we say, we can say, "All right, well, now, that's not really practical," so, let's do the next best thing, which is, let's try and optimize that testing, based on risk contribution, to try and see if we can make it more manageable.

We're giving something up by doing that, but it may be one of these cases where we don't have an alternative.

But theoretically, up to this point, we could go ahead and do the testing, to come up with the actual failure probability, in all context.

MR. CHU: The second risk-informed idea is say, in certain part of PRA, you need safety injection and you need the actuation signal, and you can possibly look at the failure probability of the safety injection system, itself, and if it has a failure probability of 10 to the minus three, then maybe the software failure probability is there, so, we can show the software failure probability is significantly less

than that, then probably the software failure probability is low enough.

In that sense, maybe we don't have to do the --

MR. KURITZKY: Keeping in mind, dependencies, which might --

CHAIRMAN STETKAR: Yes, I was going to say, I read that in the report. Boy, I would advocate against that second notion.

We've had ample experience with trying to do that in other areas, and the problem with the -- you know, I recognize, you've constrained yourself down to only does the reactor trip or not trip?

I like to think of safeguard actuations systems, which are much more pervasive, in terms of the things that they affect.

So, what are you going to do there? I mean, is it injection? Is it a combination of injection and recirculation? Is it a combination of, you know, the hardware, you know, auxiliary feedwater, all of those functions are controlled by the ESFAS system --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- and if you start down a path of comparing -- you know, is it the

software failure for the auxiliary feedwater function, only, that you're comparing numerically? The plant doesn't work that way.

So, you know, I can understand it, in a reactor trip sense, that as long as you can assure yourself that the software contribution is smaller than the failure of the reactor trip circuit breakers, that's pretty easy.

But you're only looking at a very simple bimodal function, again.

I think that the construct, you know, up until that second bullet, you're proposing -- I don't see why it shouldn't work for a broader sense of like safeguards actuation, which gets out into many related -- some related, some not so related, except through the safeguards actuation system, itself, functions.

MR. KURITZKY: Right, and I think --

CHAIRMAN STETKAR: So, this comparison with the hardware is, like I said, it might work for reactor trip breakers, but not so clear for other things.

MR. KURITZKY: Right, I think in this case, it would have to be in comparison to the hardware associated with the safety injection actuation.

In other words, if failure --

CHAIRMAN STETKAR: But if there isn't any, I mean, you know, if that hardware is finally distributed among end-state real, you know, end-state devices that tell valves open and closed, and pumps to come on and off, I think it's really, really difficult to try to infer what the acceptable hardware failure -- you know, what is it that you're comparing against?

MR. KURITZKY: Right, what is the failure probability of the hardware?

The only thing that could work, is it's compared to the --

CHAIRMAN STETKAR: It would have to be the hardware itself, yes --

MR. KURITZKY: It has to be the part that's in common to all the -- where all the signals are going. It can't be, once it goes downstream, it splits.

CHAIRMAN STETKAR: Okay, okay.

MR. KURITZKY: Yes, I don't think we're going to be relying on that one.

CHAIRMAN STETKAR: I had a side comment on that second bullet, and it's a -- I wouldn't spend a lot of effort going in that direction. That's just me.

MR. KURITZKY: I've noted it, I agree.

MR. CHU: In terms of generating test cases, here, as we probably already discussed most of the bullets, we use the context defined in PRA.

For example, for each initiating event, we define a context for the RPS, and for each event tree sequence that include an function, defines a context for ESFAS, and for each context, we need to generate test basis and estimate software failure probability.

So, generally, in the test cases, we consider other things that can vary generating, such that it can possibly generate different input, and some of the input can potentially trigger a failure of the system.

Some of the -- we can look at it as part of the operational profile definition. For example, we can consider the initial condition of the plant, and in case of LOCA, we can consider different LOCA size and LOCA locations.

In addition, we can look at inaccuracies or the time delay of the sensors. These are the additional variations that form the operational profile for the context.

CHAIRMAN STETKAR: I'm assuming that when you talk about time delay of the sensors, you'll allow

them to vary outside of their nominal design basis conditions.

This gets back into other off-normal conditions that are, you know, have occurred in practice, not very frequently, but where sensors fail in certain ways that weren't predicted by, you know, the acceptance criteria for the sensor.

MR. KURITZKY: And I don't think we've actually gone through that level of detail, yet.

CHAIRMAN STETKAR: No, I understand.

MR. KURITZKY: That is something we want to make sure we include, right.

CHAIRMAN STETKAR: Yes, I mean, it just comes back to that.

MR. KURITZKY: Yes.

CHAIRMAN STETKAR: It's trying to capture what the range of that operational profile might be.

MR. CHU: And then, one thing I've seen in the literature is, people, when they do test, they emphasize the importance of an article. It's like looking at your test output and determining the output, if the output is successful.

Since we are looking at protection systems, I think maybe the article will be simple. That is, they input other test cases, required

actuation signal be generated, therefore, we just look at the -- when we do the test, we look to see if the actuation signal is generated.

Therefore, maybe the article is not a big issue for what we --

CHAIRMAN STETKAR: It's certainly much simpler in this particular context, then trying to judge the quality of a control function.

You know, do you get enough feedwater flow for this particular context, where enough is not a well defined quantity.

MR. KURITZKY: Right, and it varies a lot.

CHAIRMAN STETKAR: And it varies a lot.

MR. KURITZKY: And in this particular case, one thing, too, is that, of course, we might have to consider a little bit about the timing. You know, the signal has to be generated within a certain time frame.

CHAIRMAN STETKAR: Right.

MR. CHU: This slide proposes to do a feasibility study, by estimating the total number of tests needed to demonstrate that the software failures contribution is low enough.

There is a study that we can do, without performing any tests. Basically, they take a PRA

model and use the PRA model to carry out this feasibility study.

We start with the PRA model, given the PRA model, Reg Guide 1.174, following Reg Guide 1.174, we can determine what is an increase in CDF that's considered acceptable, and effectively, we allocate this increase to different sequences, in which software failure can occur.

We can allocate it in an optimized way, such that the total increase is within the limit specified by Reg Guide 1.174.

So, the work will include, identify all sequences involving protection function failure and then, ideally, if we think of a plant wants to replace the analog system with a digital system, ideally, we will hope the digital system is more reliable than the analog system.

Therefore, if we have a model for the digital system and estimate the hardware failure of the digital system, and it's failure probability is smaller than that of the analog system, then the difference in the failure probability can be allocated to software failure, such that the net -- there is no net increase in the total CDF.

This is one way of looking at it. Of

course, at this point, it's not a commonly accepted model for quantifying digital hardware failure. Therefore, mostly likely, we can use just the PRA with the analog system, and try to use the Reg Guide guidance that is -- allocate the total -- the allowed increase in CDF to different sequences in a somewhat optimized way, such that some of those frequency sequences, or those low frequency sequences, the software reliability or failure probability doesn't have to be very low, because the sequence do not -- to start with, do not contribute much to the total CDF.

So, this is the basic thought of the feasibility study, somehow, optimally allocated delta CDF to different sequences, such that we can reduce the total number of tests, and the total number of tests kind of gives us a measure of how difficult it is to do a thorough job.

And so, this is something that typical PRA guy can carry out, without actually doing the tests. Of course, it depends on the specific plant PRA, as I indicated earlier.

If we look at RPS and if a plants ATWS sequence are not important, then probably, prior software failure probability is still acceptable.

MR. KURITZKY: It also gets to your

previous example about the multiple two plants with the different number of diesels.

A plant that has a very large ATWS contribution would not do very well in that regard, whereas a plant that has a very low ATWS contribution would have, theoretically, more margin to play with.

But again, it's not optimal, because it's not coming up with the software failure probability, it's just --

CHAIRMAN STETKAR: I keep coming back to, you know, this whole focus seems to have changed the fundamental, what I thought, was the fundamental direction of this research, which is, how do we define software failure modes and estimate the reliability of the software within those failure modes?

We're now saying, well, we might be able to develop methods, if I'm upgrading my analog reactor protection system, to a digital system, ways that people might be able to convince the NRC that, well, I can convince you that my software reliability is not worse than 'x', and therefore, that's an acceptable risk-informed change under 1.174, which doesn't help me an awful lot, if I'm a new plant and I'm estimating a total core damage frequency of 10 to the minus eight, and I don't even have a Reg Guide 1.174 set of

risk metrics that may apply in there, and I'm not quite sure how that's going to be done and carried forward.

MR. KURITZKY: Right, though, even with the risk-informed tech spec change for a new plant, and again, that process, regardless of digital I&C, I don't think has necessarily worked out, as to what would be an acceptable change.

CHAIRMAN STETKAR: Right.

MR. KURITZKY: Whatever process would be used to determine what's an acceptable change, maybe someone comes up and says, "Ten percent increase is okay, we'll let you make the change," but that same 10 percent, we could theoretically use in this approach, also.

But it's a difficult thing. It's not nearly as useful and flexible as what we also were hoping to get.

The reality of it, as we looked into these various approaches, there was no approach that looked like an easy answer, and we're trying to see what is the best we can get out of it, with what we hope to be, at least, somewhat practical application.

Meanwhile, I sit by the phone, waiting for the call that tells me, "Hey, you forgot about this

method that will solve all your problems," but the phone hasn't rang yet.

CHAIRMAN STETKAR: No, that's a valid, you know, a valid comment. It's just a matter of making it very clear, what the research program is producing --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- and how that may deviate from the originally stated intent of the research, and because of what it is producing, what the limitations on someone's use of that information may be, or the methods, or whatever.

MR. CHU: I think this slide summarizes the key assumptions and the limitations. This relates to John's question, or John's issue, on he seemed to be interested in estimating the absolute value of software failure probability.

So, these are -- these bullets kind of are limitations on the statistical testing. For example, whether or not the operational profile truly represents the real condition, and if it is not 100 percent representative of the true conditions, by how you're going to miss when you estimate the failure probability?

Similarly, the realism of the test

configuration, if you don't use realistic test configuration, by how much do you miss, when you estimate the software failure probability?

The memory issue, we have talked earlier, is another issue that place limitations of the accuracy of statistical testing.

In addition, we want to point out the internal failure of the digital system itself can contribute to it and can challenge the software, and typical statistical testing doesn't assume internal failure, so that's another factor that can affect the accuracy.

DR. GUARRO: Can you clarify --

MR. CHU: The estimate is the large number of tests needed.

DR. GUARRO: Can you clarify this last statement you made? Statistical failure would not -- I mean, it could, it could account for internal failures of the digital system. That's just another scenario that you'll have to assume.

MR. KURITZKY: Well, that's another scenario. That's a whole other large set of scenarios, right, yes, exactly, and that's one more thing that can --

DR. GUARRO: I use the word scenario in

quotes.

MR. KURITZKY: Right, that's just one more thing that defines the operational profile. That would be just be one more factor. We can talk about LOCA location and size and you can talk about sensor values and thermal hydraulic parameter, and all the different possible combinations of digital system hardware failures.

Quite honestly, that might be more important to the -- affecting the software failure probability than many of the other things that we might vary, and I don't know whether or not, you know, that is something that is practical to vary or reasonable to vary, or not.

You know, it's just one more --

DR. GUARRO: Yes, I agree with everything that you're saying. I'm just saying that that bullet is, you know, kind of says that it's -- it implies that it's not possible to approach the issue.

MR. KURITZKY: Well, I think that --

CHAIRMAN STETKAR: I don't think it says it's not possible.

MR. KURITZKY: It's just that we're not --

CHAIRMAN STETKAR: It's just a statement of fact.

MR. KURITZKY: Right, it's just that our approach --

DR. GUARRO: No, that's an inaccurate fact, because --

MR. KURITZKY: Well, no, no, in our approach --

DR. GUARRO: You know, there are testings that, you know, I put aside failures inside the controller itself, you know, and you know, full tolerance of the software is one of the issues, one of the ways that people -- the design feature that people have to try to counter that. So, it is part of the whole, you know, ball of wax.

DR. HECHT: Why don't you say, in this study?

MR. KURITZKY: Yes, that's what I was going to say, I was going --

CHAIRMAN STETKAR: Okay, so, it's --

MR. KURITZKY: I have to agree with Sergio, it is a little misleading, because if these are the key assumptions of the method that we're using, then that's just a statement of fact.

If this is key assumptions of statistical --

DR. GUARRO: Just the general --

MR. KURITZKY: -- then that doesn't necessarily have to be the case, because someone can --

DR. GUARRO: Yes, I'd rather the second way, that's why I was --

MR. KURITZKY: Right, and I agree, it's misleading, right, and to tell you the truth, I'm not sure whether or not that's something we should necessarily rule out, and even in our study, but I'm sure practical limitations are going to make us rule out a lot of things that we're going to wish we could have included.

MR. CHU: Going back to the desirable characteristics, I think our approach, proposed approach, somewhat better addressed three of the characteristics.

The proposed approach certainly takes into the context, and if our approach is applied successfully, you can think of it as not a validation of the method, but these -- it's a demonstration of the method, successful demonstration of the method.

By considering risk information maybe, we can show software failure probability does not have to be very low.

MR. KURITZKY: Just to clarify that last

slide, to make sure what -- so, those were three characteristics that didn't get 'yes' in the original evaluation of statistical black box testing, but that might be closer to 'yes' with this study.

CHAIRMAN STETKAR: Okay.

MR. CHU: This slide talks about the new operating control system that we just -- example system, we cannot apply statistical testing to, and we —- Advanced Test Reactor has five in-pile tubes that are equipped with pressurized water cooling loops, and this LOCS is the control system of the cooling system, and it also has some protection function. That's the part that we are trying to test.

That is, it will detect some of the normal condition in the loop cooling system, and generate reactor trip signal.

CHAIRMAN STETKAR: I hesitate to ask this, but I guess I have to.

Is there a PRA available of this facility?

MR. KURITZKY: Do we know, yet?

MR. CHU: I think they do have a PRA. We are trying to get more information about it, and then as far as I know, this cooling system has a thermal hydraulic model of it, and that thermal hydraulic model is said to be used to support some kind of

design basis analysis.

So, in that sense, I kind of expect in the PRA, they will include some kind of accident scenario, associated with this cooling system.

MEMBER BLEY: Is that is what is driving you to the ATR?

CHAIRMAN STETKAR: That is the only thing that they could get any software --

MEMBER BLEY: Okay.

CHAIRMAN STETKAR: This is the only thing that they get a --

MR. KURITZKY: Right, it was the last first --

CHAIRMAN STETKAR: And actually --

MEMBER BLEY: Fair enough.

CHAIRMAN STETKAR: Well, the question about the PRA, obviously, is valid, because it's not clear, if I pick up the PRA, you know, the SPAR model for plant 'x', operating nuclear power plant, and somehow infer that sequences through that can be used to generate, you know, the testing for this particular thing, which you happen to be able to test, in the test case --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- you know, I'm

starting to feel a little bit uneasy about patching together, you know, a Frankenstein.

MR. KURITZKY: Right, I understand. I understand, yes, I think our hope is that there is a PRA that we can get access to, that will give us enough information to establish the context.

CHAIRMAN STETKAR: Well, we'll worry about, if the answer is, "No, they don't have one," then --

MR. KURITZKY: Then Frankenstein is the --

CHAIRMAN STETKAR: Well, but I mean, that is a point, in terms of a feasibility study, demonstrating a methodology -- just different parts of that --

MEMBER BLEY: If you Google it, advanced test reactor, safety improvements, from the PRA --

CHAIRMAN STETKAR: Okay, so, that must have been a -- that is really important -- okay.

MEMBER BLEY: And there is some seismic PRA stuff in here, too.

CHAIRMAN STETKAR: Okay, but the appropriate three letters PRA appear for that, good.

MR. CHU: So, in our case study, we're going to try to demonstrate how statistical testing can be performed to estimate software failure

probability, or possibly an upper bound of software failure probability, and what we plan to do is perform thermal hydraulic calculations for the scenarios associated with this cooling system that generate the test cases, and the time history of the physical parameters can be saved in a file.

Then the test case information, the file can be fed to the software, and then observe the output of the software to see if the actuation signal is generated.

So, this is the part where there is a lot of uncertainty. We don't know what's available to carry out this kind of testing.

We're getting more information. We know that they have some simulator associated with this loop cooling system, and I believe they use it to do some kind of -- they were able to use it, to do some kind of testing of the software.

Therefore, in order to -- we can possibly try to use the simulator, but somehow, some kind of interface software will be needed, such that we can feed the test cases to the simulator, and then observe the outcome, to see if the actual signal is generated.

This is how we are -- we are still working with Idaho Labs, to try to find out more about their

thermal hydraulic code, their simulator, to see if this third bullet, how the third bullet can be accomplished.

I think we certainly need to get more information, but so far, this is the bullet that has pretty large uncertainty, and they have a PRA, and hopefully the PRA covers scenarios associated with the loop cooling system. At least it will give us, you know, frequencies of sequences associates with this system, and it can help us formulate the statistical testing approach.

I think this is the last slide.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: Yes, last slide.

CHAIRMAN STETKAR: Okay, there is a lot there.

MR. KURITZKY: Yes, there is. Okay, so, if there is no further questions for Louis, I just have a couple of slides for wrap up.

Okay, just to wrap up, you've heard a lot today, over quite a number of hours, about what we propose to do.

You've seen our grand ambitions and you've also heard quite extensively about all the pitfalls that we are sure to have to deal with.

MEMBER BLEY: Not to throw you off, before you begin, but the one thing I didn't see mentioned, you had the workshop a couple of years ago, to kind of see what is out there and get direction.

Are you -- do you have any scheduled peer review of any kind, on this work with leading experts in this area?

MR. KURITZKY: Well, yes, I mentioned before, we've had, every stage of the way, with this project, we've tried to have a substantial amount of peer review.

In every report, we've had a targeted excellent peer review panel that we put together, to comment. All of the NUREGs, we've sent out for public comment, as well.

All the reports, both the technical reports as well as the NUREGs, we've sent to NRR, NRO, both the engineering and PRA groups, the digital I&C and the PRA groups, to get their comment on.

So, we've done -- and we've also presented a lot of the stuff at conferences, too, both in the PRA realm, as well as the software reliability realm, to try and get as much feedback and input from various sources, as we can.

So, yes, we've tried to get -- taken the

extra step to try and get as much peer review and external input as possible, in all this work.

The specific workshop that we had in 2009, up at BNL, with that specific set of experts, the focus on that was establishing the philosophical basis from modeling software in a probabilistic sense.

Because we had everybody there, we wanted to take advantage of that. So, after getting that first item out of the way, we also explored with them, ways of modeling software for use in the PRA, as well as possible ways to quantify it, and some of that, Louis had mentioned in his presentation, how it was -- the methods we came up with were consistent with what was identified by that group, you know, the BBN approaches and statistical testing. They were both listed by that group, when we met with them.

Several members of that group have also been retained -- have been used subsequently as peer reviewers for some of the later reports, and some of those people, in fact, are working with us on the current work.

So, yes, we've tried to continue to make use of that expert -- you know, that body of experts.

CHAIRMAN STETKAR: Alan, again, before you get rolling on this the next couple of years, what is

the status of this current draft NUREG? Is it out for public comment already?

MR. KURITZKY: Yes, this --

CHAIRMAN STETKAR: And what is the end of the public comment period?

MR. KURITZKY: It went out for public comment on May 18th.

CHAIRMAN STETKAR: Okay.

MR. KURITZKY: Public comment period ends on, I think it's like July 17th or 18th, something like that, and then we're going to go ahead and we're going to take all the external peer review comments, the public comments, feedback from the Subcommittee.

Again, like I've said, we sent it over to NRR and NRO, both the PRA and the engineering groups there, and to the -- and research the engineering group, the digital I&C branch.

So, we're going to take all that feedback together, and we hope to have a draft final report, it's going to go through extensive writing an rewriting, but we hope to get a draft final report back from BNL, sometime late in the year, but it will be quite a while until it gets actually published, because there is going to be a whole bunch of --

CHAIRMAN STETKAR: I was going to say, you

know, there has been quite a bit of discussion today, and I think you get a sense of somewhere between cautious optimism and skepticism, in terms of how well this process may work, and what information we get out of it.

And also, this potential hope that indeed, the voice will come out of the woodwork and say, "Yay, I have something that may work better than this."

MR. KURITZKY: Right.

CHAIRMAN STETKAR: Will we have an opportunity to see what those public comments are, and how you, you know, addressed them, you know, before the final NUREG is issued, because I think, at least, I would, and I suspect other members might be interested in seeing the results of that feedback.

MR. KURITZKY: Yes, you know, right up to now, with the reports that --

CHAIRMAN STETKAR: I'm thinking of scheduling, you know, future interactions --

MR. KURITZKY: Right, and then the last slide I have talks about the schedule --

CHAIRMAN STETKAR: But I looked ahead, that's mostly on --

MR. KURITZKY: Right.

CHAIRMAN STETKAR: -- presuming you

progress full steam ahead on your --

MR. KURITZKY: Right, and what I was going to mention is, in the past, we've actually scheduled these briefings, when the report has gone out first available, the draft has been available publicly.

CHAIRMAN STETKAR: Right.

MR. KURITZKY: Tried to be as close to that as possible, so we have enough time to factor Subcommittee input before we do the final, but we haven't generally come back after.

All we've done is, when the next draft has shown up, we've talked about the previous file, but it hasn't -- but that doesn't give the Subcommittee a chance to weigh-in, prior to that, you know, it's usually after that file has already gone out the door.

CHAIRMAN STETKAR: Right, in this particular, I don't know, Dennis, what do you think?

Should we have an interaction, after they -- you know, if they get public comments back in July time frame, in sort of the September time frame --

MEMBER BLEY: Well, they have scheduled one in September for us.

MEMBER SIEBER: They have.

MEMBER BLEY: It says here on the fourth slide, you're coming back to us in September.

MR. KURITZKY: Wrong year.

MEMBER BLEY: Wrong year? Oh, that's next year.

CHAIRMAN STETKAR: That is September 2012, to give us feedback on the --

MEMBER BLEY: After the draft on the --

MR. KURITZKY: Next year.

CHAIRMAN STETKAR: Draft on the case study.

MR. KURITZKY: Right, the next report. Yes, you know, after --

CHAIRMAN STETKAR: The draft on what they've done on the case study.

MEMBER BLEY: Oh, okay.

CHAIRMAN STETKAR: Presuming that is --

MEMBER BLEY: I thought that was the combined report which was methodology with an application.

MR. KURITZKY: It would just be a separate report on the application.

MEMBER BLEY: Okay, so, then, yes.

CHAIRMAN STETKAR: Okay, thanks.

MEMBER BLEY: I misread your slide ahead of time.

MR. KURITZKY: No problem.

CHAIRMAN STETKAR: That's why I wanted to intercept it, before we got to that next slide.

MR. KURITZKY: But the thing is, schedule wise, when we get those comments back in July, the process, we typically get a lot of comments, because it goes to so many people.

So, it's going to be months before we have --

CHAIRMAN STETKAR: Yes, well, I didn't want to -- September is the date.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: I just wanted to set the notion, that before that, we do want to have an interaction sometime between the point that you process the public comments and the time that you go out the door with the final version of the NUREG.

MR. KURITZKY: Okay.

CHAIRMAN STETKAR: Just so that we understand --

MEMBER BLEY: When are you aiming for the NUREG to come out?

CHAIRMAN STETKAR: Sort of the end of the year.

MR. KURITZKY: Well, no, that's just to get a draft. That's just to get -- it's a long time,

once we get the draft in, from BNL, they're going to go through it and address all the comments, and adjust, incorporate them into that --

MEMBER BLEY: I think that is the time we'd want to --

MR. KURITZKY: Right, but what happens is, it comes to me, and there is a good two month back and forth, before I'm comfortable with everything and we readjust things.

MEMBER BLEY: Yes.

MR. KURITZKY: Then it starts to go through my management review. So, it's a pretty long process and when we come here, it's going to have to be after I get a chance to work with BNL on it, and I'm not sure, you know, there may be a certain extent of management review that may be required, before we're allowed to provide it, also.

So, we would have to figure, I could go back and talk to management --

CHAIRMAN STETKAR: I mean, the message, I think, you know, regardless of the specific months and weeks, and all that kind of stuff, is that we would like an opportunity to be briefed from you, after you've had gone through whatever exercises, to process the feedback, from the public comments, and

disposition that feedback, and with ample time before the NUREG is scheduled to go out the door, final, such that we can understand what the nature of the feedback was, what comments you've received from the public, and any other reviews that have been done, and how you've responded to those.

MR. KURITZKY: Yes, okay.

CHAIRMAN STETKAR: On this particular NUREG, regardless of whatever the schedule is on development of reports for the --

MR. KURITZKY: The case studies.

CHAIRMAN STETKAR: -- the path forward on the case studies.

MR. KURITZKY: Right, yes, we will --

CHAIRMAN STETKAR: And you know, scheduling is obviously an issue, but that's a notion of, I think we'd really like to hear back from you on that.

MR. KURITZKY: Okay, like I said, I'll bring that back when I discuss back with management, which we've got to schedule for when we can work it in.

Okay, so, clearly, as it should be evident by now, the current focus of our PRA digital I&C peer review research is getting software failures into the

models. That is what we're working on right now.

There is, of course, a lot of related work that's being done in other areas of the NRC, the digital I&C branch in research is addressing a number of different things, and you're going to hear about some of those in just a couple of weeks.

But the main focus of our work, as we discussed today, was identifying a couple of methods that we could try out in proof of concept studies for quantifying software failure probability.

We discussed our review of the various methods out there, a set of desirable characteristics, which we compared them to. We've come up with the BBN and the statistical testing as two preferred methods that we would like to try out.

Again, I emphasize that that doesn't mean the other methods do not have potential, but these are the two for our various reasons, that I think that Meng and Louis have described, that we wanted to pursue, at this point.

Again, because of all of the limitations that we have identified today, what we want to do is use the BBN approach to come up with a prior distribution and then use the statistical testing results to update that standard Bayesian update

process.

We also are looking into the possibility of using risk-informed testing to try and minimize the test burden, and then incorporate contexts into the software, but that's obviously -- each successive improvement of the process entails a much greater resource and complexity to it, also.

So, we're just going to have to dive in, and see, you know, how far we can get.

Again, as we mentioned previously, we're going to multiple -- the BBN model, the amount of effort it takes to do a detailed BBN model is fairly extensive. So, we don't want to necessarily just do that, and even if it works, say, "Okay, everybody, you now must do a very detailed BBN model."

We want to look at a simplified, more high level also, see how well that approach works. As was discussed earlier, that even may give you a better, you know, a better job at estimating the failure probability than the detailed model.

A third one we want to do also is, just use a non-informed prior and use statistical testing data on that.

Quite honestly, going back to what Dr. Hecht mentioned before, you know, if everybody was to

have the same -- if all the nuclear software is of a similar type, in terms of its software life cycle design activities, et cetera, quality of those activities, you may expect very similar results out of the BBN, maybe you don't need to actually do the BBN. You could use some non-informed prior and just use the statistical testing that may drive the answer.

Even if they're all the same, it may still be important to account for the software life cycle activities, and therefore, use the BBN, but we want to also just use the uninformed prior, just to try and see, as another standard for comparison, just how it lines with the two BBN models.

I already mentioned the fact that we were going to try and do the risk-informed testing to the extent possible.

One thing I forgot to mention earlier, it was on my earlier slides, in my first presentation, but I forgot to mention it, was the Korean Atomic Energy Research Institute, KAERI, they were involved with one of our peer reviews.

You can see from some of the earlier reports, they've done quite a bit of work in this area. They have a lot of experience in applying digital systems into PRAs, because they have them with

their reactors, and also, they particularly have experience in applying the BBN approach to digital systems software.

So, go ahead.

CHAIRMAN STETKAR: Let me ask about that, I see the last bullet, I hear what you're saying.

The case studies, however, will be done in the context of the ATR cooling system and the PRA.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: There is no problem with sharing that information --

MR. KURITZKY: With Idaho, we've got --

CHAIRMAN STETKAR: Okay, fine.

MR. KURITZKY: Up to now, all the information we've been provided, they've told us is public information, can be shared with the public.

Now, I don't know, again, as we continue to try and get more information, because as I think Louis and Meng both mentioned, there is more information we really would like to have from them.

Now, you know --

CHAIRMAN STETKAR: Well, the reason I ask that is, if you do run into a difficult with sharing the information, how critical is the cooperation of KAERI, you know, for your projected successful

completion?

MR. KURITZKY: Right, right, I think we benefit a lot by having KAERI involved. It's a good leveraging of assets. They're providing a lot of manpower to the effort and experience and work, and I think that what we would happen, if we did end up in a situation where some information was not -- we couldn't share with them, I think we would still try, to the extent possible, use them for aspects of the study that didn't require them to see that information.

CHAIRMAN STETKAR: Can I flip it and say, since the Koreans have done an awful lot of work in this area, do they have actual digital systems and PRAs of those systems available, that they would share in a collaborative way, with you?

MR. KURITZKY: In the very long drawn out process, and then trying to identify an example system for these case studies, it's gone over months and months and months, that was definitely one that was on the board.

They have what is called the KNIC system, Korea Nuclear Instrumentation Control System, which KAERI themselves designed, and they carried that forward, that system development, up to a certain

point, at which point, they then gave it to a commercial vendor to finish up the work on that.

So, they are intimately familiar with that system, and so, that was one system that we thought might be of value in doing the case study.

The problem with that system is, most of the documentation is in Korean, which doesn't bother -

CHAIRMAN STETKAR: In Korean?

MR. KURITZKY: Yes, but anyway, it makes it a little more difficult on our end of the study, and so, there are logistical issues.

But up until we got the Idaho system, that was still on the back burner, as something, if the Idaho system didn't pan out, we would go with the --

CHAIRMAN STETKAR: So, there is a potential, what I'm hearing is, there might be a plan B, if the information from Idaho can't be shared?

MR. KURITZKY: But we're hoping, of course, not to have to go to plan B, because like I said, there is a lot of -- Louis, do you want to mention something?

MR. CHU: There is already one document, currently, that is limited to U.S. citizen, only. Idaho prepared description of the system, 30 to 40

pages, and they classify it as the --

MR. KURITZKY: "No foreign".

MR. CHU: -- may contain, I don't know the exact words, but it's control --

MEMBER RYAN: Sensitive uncontrolled information, SUNSI, probably.

CHAIRMAN STETKAR: There is a 'no foreign' classification.

MR. CHU: So, we're in the process of creating a public version of it, by removing some more sensitive information.

CHAIRMAN STETKAR: If there is a PRA available, out there, and it sounds like there is, I'd contact them about that, because people tend to get awfully protective of PRA information.

MR. KURITZKY: Now a days. Post 9/11, they didn't.

CHAIRMAN STETKAR: Post 9/11, yes.

MR. KURITZKY: That is a good point, Louis, right, when you check with them on the PRA, ask them also about whether that is something that we're able to share with the Koreans.

MR. CHU: Yes.

CHAIRMAN STETKAR: Because without that, establishing those contacts, for at least that part of

the process --

MR. KURITZKY: Right, but to go back to what I said before, we would still probably want to use the Koreans to the extent we could. Most of their experience is with the BBN approach, and I think we could help set that up, without even having seen the PRAs.

CHAIRMAN STETKAR: Yes, okay.

MR. KURITZKY: So, we still will benefit from that collaboration.

Okay, so, I think that was the one we had before the interactions.

Okay, so, we discussed this a little bit, and you guys went ahead.

The two cases studies, we're going to have NUREG/CRs on those and the draft for those are tentatively going to be done in August 2012 and February 2013, and then there is several months for getting the public comments and preparing draft.

So, typically for those reports, we would want to come to the Subcommittee after it's publicly available, but as soon afterward as possible, to allow time to address things, as we incorporate everybody else's comments.

But of course, it's up to you guys to

decide, when you would like us to come, and both of those cases, too, we would not be requesting a letter. It would just be information exchange and getting feedback.

But going back to what you said before, Dr. Stetkar, you know, we can certainly try and schedule something after we do the incorporation of public comments for this current draft NUREG.

CHAIRMAN STETKAR: I think, you know, to reiterate, I think we'd be interested in that, and I think we need to keep a dialogue open because of the time frames that you're talking about here, out over the next 18 to 24 months.

MR. KURITZKY: Right.

CHAIRMAN STETKAR: We might be interested in an update, in some sort of interim time, about how you're doing, sort of a briefing, you know, not a full day, but maybe a half-day briefing on how things are going.

MR. KURITZKY: Right, now, the only thing --

CHAIRMAN STETKAR: Rather than waiting until, you know, the draft NUREG hits the street.

MR. KURITZKY: Right, the only thing that kind of limits, and I understand these time periods

are pretty drawn out, and so, it doesn't give us a fairly frequent interaction, but kind of an unwritten rule was that we were supposed to have some kind of document to be able to give before we come back --

CHAIRMAN STETKAR: Yes, that is -- I recognize that, and --

MR. KURITZKY: But I did come one time without a document, I think actually, I was here last August and --

CHAIRMAN STETKAR: We've had some kind of informative briefings in the past, that -- without documents. We, in general, do request documentation.

So, let's just keep that --

MR. KURITZKY: Right, in mind, and again, it's up to you. If you want us to come, even though we don't have a document, just to give you a briefing, we can make slides that kind of cover all the work that's being done, and provide that to you ahead of time, and then come talk to that, and then that's fine.

CHAIRMAN STETKAR: We've had those types of briefings. They're rare, but we've had those types of briefings in the past, especially where something is as -- it's called as dynamic as this might become over the next 18 to 24 months.

MR. KURITZKY: Okay, so, that's pretty much all that we have to say.

I mean, we certainly appreciate the time and input. We got a lot of feedback today, that we'll take back and consider, because there is a number of important issues that got raised.

Of course, some of them, we were aware of. Some of them, new takes on it that we want to certainly consider, so, we appreciate that, and are there questions?

CHAIRMAN STETKAR: I'd like to thank you. I think we had an excellent interchange. It's always useful to have these exchanges, from I think, both ways.

Before we close, I guess, I'll go around the table, as we do traditionally, and make sure that we've captured all the questions or concerns that anybody might have. Jack?

MEMBER SIEBER: No additional questions. Thank you.

CHAIRMAN STETKAR: Dennis?

MEMBER BLEY: Nothing additional. I think I threw everything on the table.

CHAIRMAN STETKAR: Mike?

MEMBER RYAN: I just want to thank the

presenters and the members and everybody at the table. It was an interesting interaction.

It's not my area of principle expertise, so, I learned a lot, but thank you so much for your hard work and your excellent work in presenting it today. Thank you.

CHAIRMAN STETKAR: Sergio, anything final?

DR. GUARRO: No, I'll be uncharacteristically silent.

DR. HECHT: I just wanted to say that what I call allocations, what you called the change in direction to the upper methods, I think it's a good way to go to help.

CHAIRMAN STETKAR: And again, I don't have anything. I'd like to thank you, again, and appreciate all the time and effort. You put a lot of time and effort, obviously, into this presentation and the work, and we're really interested in following the path forward. So, thanks, again.

One last thing. I don't believe we have the bridge line open anymore?

MS. ANTONESCU: No, we don't have that, apparently we don't, no.

CHAIRMAN STETKAR: And I just need to see, anybody from the public here wish to say anything?

(No response.)

CHAIRMAN STETKAR: Thank you.

MEMBER BLEY: We wore them out.

(Laughter.)

CHAIRMAN STETKAR: If nothing else, it's a strategy. We are adjourned.

(Whereupon, the above-entitled matter concluded at 4:41 p.m.)

United States Nuclear Regulatory Commission

Protecting People and the Environment

Overview of Digital I&C PRA Research

Advisory Committee on Reactor Safeguards
Digital Instrumentation and Control Subcommittee

June 7, 2011

Alan Kuritzky
Division of Risk Analysis
Office of Nuclear Regulatory Research
(301-251-7587, Alan.Kuritzky@nrc.gov)

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Purpose of Briefing

- Keep the Subcommittee informed of staff progress in the area of digital instrumentation and control (DI&C) and software probabilistic risk assessment (PRA)
- Obtain feedback from the Subcommittee on the proposed approach
  - Selection of quantitative software reliability methods (QSRMs)
  - Bayesian Belief Network (BBN) method
  - Statistical testing method
- No letter requested

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Background

- Current licensing process for digital systems is based on deterministic engineering criteria
- Commission's 1995 PRA policy statement encourages use of PRA to the extent supported by the state-of-the-art
- Risk-informed analysis process for DI&C systems has not yet been fully developed

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Objective

- Identify and/or develop methods, analytical tools, and regulatory guidance for:
  - Including digital system models into nuclear power plant (NPP) PRAs
  - Incorporating digital systems in the NRC's risk-informed licensing and oversight activities

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Previous Research (1 of 2)

- Failure mode identification and analysis
  - Brookhaven National Laboratory (BNL) – Failure modes and effects analysis of a digital feedwater control system (DFWCS) (NUREG/CR-6962 [2008], NUREG/CR-6997 [2009])
- Initial reliability modeling
  - Ohio State University/ASCA/University of Virginia – Dynamic reliability modeling methods applied to a DFWCS (NUREG/CR-6901 [2006], NUREG/CR-6942 [2007], NUREG/CR-6985 [2009])
  - BNL – Traditional reliability modeling methods applied to a DFWCS (NUREG/CR-6962 [2008], NUREG/CR-6997 [2009])

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Previous Research (2 of 2)

- Operating experience analysis
  - BNL
    - Study of reliability methods and data used by non-nuclear industries (internal letter report)
    - Collection of failure data and development of database for probabilistic modeling of digital systems (internal letter report)
    - Review of software-induced failure experience (Appendix C to initial draft version of NUREG/CR-6962; modified version will be made publicly available in 2011)
  - Oak Ridge National Laboratory (ORNL)
    - Review of operational experience data to identify generic DI&C system failure modes and failure mechanisms and to obtain generic insights (Publicly available ORNL report – ADAMS ML102210520)

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Current and Near-Term Activities (1)

- NRC/BNL currently pursuing incorporating software failure into digital system reliability models
  - Workshop on philosophical basis (ADAMS ML092780607)
  - Review of quantitative software reliability methods (ADAMS ML102240566)
  - **Development of two technically sound approaches to modeling and quantifying software failures in terms of failure probability on demand**
    - **BBN method**
    - **Statistical testing method**
  - Application of approaches to an example software-based protection system in a proof-of-concept (case) study
    - Case studies will be performed in collaboration with the Korea Atomic Energy Research Institute (KAERI)

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Current and Near-Term Activities (2)

- Initiate research to address other "gaps" in the state-of-the-art
  - Failure mode identification
  - Fault tolerant features
  - Reliability parameter data
  - Dynamic interactions
  - Common cause failures
  - Human reliability analysis

Note: It is expected that reliability models of digital systems (including software) can be developed and quantified, but it is not clear whether it will be practical and useful to do so.

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

WGRisk Activity on DI&C Failure Mode Taxonomy (1)

- Objectives
  - To develop failure mode taxonomy (or taxonomies) for reliability assessment of DI&C systems for PRA
  - To provide best practice guidelines on the use of taxonomy in modeling, data collection, and quantification of DI&C reliability
- Process
  - Work to be accomplished through a series of workshops (approximately 2 per year)
  - Additional work to be performed between workshops
- Deliverables
  - Best practice guidelines on failure mode taxonomy for reliability assessment of DI&C systems for PRA

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

WGRisk Activity on DI&C Failure Mode Taxonomy (2)

- Schedule
  - First workshop held May 16-19, 2011, in Washington DC area
  - Second workshop to be held in October 2011 in Finland
  - Two other workshops in 2012
  - Final report due to Committee for the Safety of Nuclear Installations (CSNI) in 2013

U.S.NRC
UNITED STATES NUCLEAR REGULATORY COMMISSION
Protecting People and the Environment

Selection of Quantitative Software Reliability Methods

Advisory Committee on Reactor Safeguards
Digital Instrumentation and Control Subcommittee

June 7, 2011

Tsong-Lun Chu
Brookhaven National Laboratory
(631-344-2389, chu@bnl.gov)

Brookhaven National Laboratory
U.S. Department of Energy

Outline of Presentation

- Background
- Desirable characteristics of quantitative software reliability methods (QSRMs)
- Evaluation of QSRMs
- Selection of QSRMs

Background

- NRC/BNL organized and convened a workshop involving experts with knowledge of software reliability and/or NPP PRA in May 2009 (BNL-90571-2009-IR).

  Software failure is basically a deterministic process. However, because of our incomplete knowledge, (e.g., the number and nature of residual faults, and occurrence and timing of fault-triggering inputs) we are not able to fully account for and quantify all the variables that define the failure process. Therefore, we use probabilistic modeling to describe and characterize the software failure process.

- BNL developed a set of desirable characteristics of QSRMs and performed a review of such methods (BNL-94047-2010).

Development of Desirable Characteristics of QSRMs

- The desirable characteristics were developed based on the perceived need for reliability models of digital systems in a PRA and the knowledge and experience of the study team in performing research and a literature review on modeling of digital systems.
- These characteristics are used to identify promising methods for quantifying software reliability to support formulating reliability models of digital systems for use in PRA of NPPs.

Desirable Characteristics of QSRMs

1. The description of the method and its application is comprehensive and understandable.
2. The assumptions of the method have reasonable bases.
3. The method allows for consideration of the specific operating conditions of the software.
4. The method takes into consideration the quality of lifecycle activities.
5. The method makes use of available test results and operational experience.
6. The method addresses uncertainty.
7. The method has been verified and validated.
8. The method is capable of demonstrating the high reliability of a safety-critical system.
9. The method should be able to estimate parameters that can be used to account for software common cause failures (CCFs).
10. The data and needed information exist and can be collected.

Common Limitations of QSRMs

- Test profile vs. operational profile
- Context specificity
- Demonstration of high reliability
- Failure-mode-specific modeling
- Common-cause failure of two diverse digital systems

QSRMs Evaluated

- Software reliability growth methods (SRGMs)
- Bayesian belief network (BBN) methods
- Test-based methods: Black-box and white-box
- A correlation method using software development practices (Frestimate)
- Metrics-Based methods
- Standard-Based methods
- Other considerations: Context-based software risk model (CSRM), N-version programming

Evaluation of QSRMs Against Desirable Characteristics

	1	2	3	4	5	6	7	8	9	10
SRGM										
BBN										
Test-based (black box)										
Test-based (white box)										
Frestimate										
Metrics-MTTF										
Metrics- Defect Density										
Standard-based										

Legend

YES – Method conceptually meets the characteristic and its implementation is deemed practical

MAYBE – Method may conceptually meet the characteristic and/or its implementation may not be practical

NO – Method does not conceptually meet the characteristic

Selection of QSRMs

- None of the methods reviewed meets the complete set of desirable characteristics, and no single method clearly stands out as the most appropriate.
- The BBN method and the test-based methods were selected as candidates for further assessment, mainly because of the former's ability to account for the quality of the software lifecycle activities and the latter's use of standard statistical methods, including treatment of parameter uncertainties.
- This finding is consistent with the results of the expert panel meeting on modeling software failures in PRA, where the panelists specifically identified testing and BBNs as general methods with potential for quantifying software failure rates and probabilities.
- One concern with the BBN method is that it requires converting qualitative information into failure probabilities, but there may be insufficient data or experts to accomplish this.
- Concerns with the test-based methods include the availability of the equipment (hardware and software) needed to undertake the tests, and definition and characterization of the operational profile.

U.S.NRC
UNITED STATES NUCLEAR REGULATORY COMMISSION
Protecting People and the Environment

Development of a Bayesian Belief Network Model for Software Reliability Assessment

Advisory Committee on Reactor Safeguards
Digital Instrumentation and Control Subcommittee

June 7, 2011

Meng Yue
Brookhaven National Laboratory
(631-344-7140, yuemeng@bnl.gov)

Brookhaven National Laboratory
U.S. Department of Energy

Outline of Presentation

- Background
- Development of a BBN model for estimating software failure probability
- Issues and limitations of the BBN method
- Treatment of uncertainty
- Procedure for developing a BBN model of an example system

Bayesian Belief Network Models

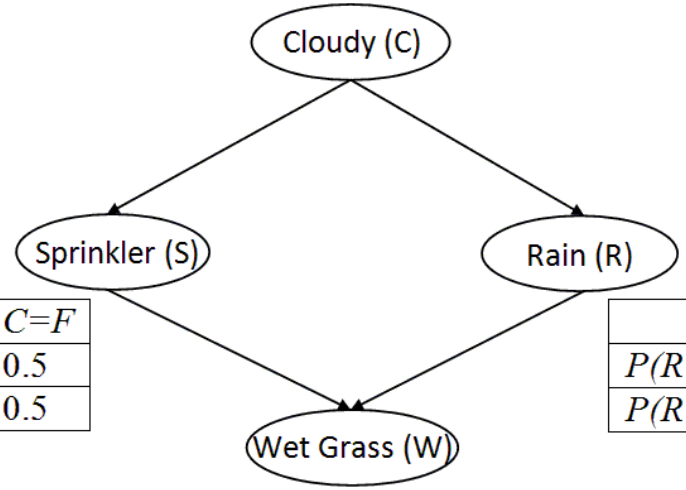
- A Bayesian Belief Network (BBN) is a probabilistic graphical model depicting a set of random variables and their conditional independencies via a directed acyclic graph.
- A basic assumption for BBN is that a node is conditionally independent of its non-descendent nodes, given its parent nodes.
- For a BBN, the joint distribution of all variables $\{V_i\}$ is

$$P(V_1, V_2, \cdots, V_n) = \prod_{i=1}^{n} P(V_i \mid parents(V_i)).$$

- Bayesian inference is performed by updating the above equation using the acquired evidence; there exists a spectrum of software tools for the inference.

An Example Bayesian Belief Network Model

$P(C=T)$	0.5
$P(C=F)$	0.5



	$C=T$	$C=F$
$P(S=T)$	0.1	0.5
$P(S=F)$	0.9	0.5

	$C=T$	$C=F$
$P(R=T)$	0.8	0.2
$P(R=F)$	0.2	0.8

	$S=T, R=T$	$S=T, R=F$	$S=F, R=T$	$S=F, R=F$
$P(W=T)$	0.99	0.9	0.9	0.0
$P(W=F)$	0.01	0.1	0.1	1.0

Reference: Murphy, K., "A Brief Introduction to Graphical Models and Bayesian Networks," 1998.

Capturing the Quality of Software Development Activities Using BBN

- Guidance on how to carry out software development activities is available in different applications, e.g.,
  - Avionic standard DO-178B
  - NUREG-0800, Chapter 7, Branch Technical Position (BTP) HICB-14.
- In general, the qualitative requirements in the standards are used to define the nodes in the BBN models which represent the quality of software development activities.
- The results of the quality assessment may be used
  - as factors that affect the addition, detection and removal of faults which, in turn, determine the number of faults remaining in the system (see, e.g., Eom [2009] and Delic [1997]) or
  - to directly estimate software failure probability by expert elicitation (see, e.g., Gran [2000]).
- After collecting information (evidence) about a specific application software development, a prior distribution of the software failure probability can be obtained using a BBN model, and the distribution can then be updated via testing and/or operational data.

Estimation of Software Failure Probability – Conversion of Fault Numbers to Failure Probability

- Software failure probability can be assessed based on the number of estimated residual faults in a software and
  - a “size distribution” of faults that was estimated from experience with similar software-development projects of the same vendor or
  - the concept of the fault-exposure ratio obtained by integrating over the size distribution of the faults from a variety of projects.
- Vendor/software-specific data are more applicable, and are preferred over the generic fault-exposure ratio.

Estimation of Software Failure Probability – Direct Estimation Using Expert Elicitation

- Expert elicitations can be used to build BBN models, including node probability tables (NPTs), to convert qualitative- to quantitative-evidence, e.g., a direct conversion of quality measures of high level BBN nodes “Quality of Product,” “Solution Complexity,” and “Quality of Analyses” to the system failure probability [Gran 2000].

P(system in failure state) (Failure_rate) (Interval Chance Node)

Quality of product	good						bad					
Solution complexity	low		medium		high		low		medium		high	
Quality of analyses	good	bad	good	bad	good	bad	good	bad	good	bad	good	bad
<0, .000001]	.01	.01	0	0	0	0	0	0	0	0	0	0
<.000001, .00001]	.7	.5	.01	.01	0	0	0	0	0	0	0	0
<.00001, .0001]	.25	.4	.7	.5	.01	.01	.01	.01	0	0	0	0
<.0001, .001]	.04	.09	.25	.4	.7	.5	.7	.5	.01	.01	0	0
<.001, .01]	0	0	.04	.09	.25	.4	.25	.4	.7	.5	.1	0
<.01, .1]	0	0	0	0	.04	.09	.04	.09	.25	.4	.8	.1
<.1, 1]	0	0	0	0	0	0	0	0	.04	.09	.1	.9

Issues and Limitations of BBN Method (1)

- Converting qualitative to quantitative information
  - A lack of data relating quality of software development to both low-level and high-level quantities, for example:
    - Relating verification and validation (V&V) quality to the detection probability of residual faults
    - Relating software development quality to software failure probability or reliability
  - Benchmarking can be undertaken to validate software reliability assessments by collecting operating experience of software projects and/or by applying statistical tests at the system and node-probability-table levels, such that qualitative and quantitative information can be related.
  - The difficulties in carrying out benchmark studies should not be under-estimated (e.g., scarcity and applicability of data to relate qualitative information to quantitative information).
  - Conceptually, the idea of regression analysis can be extended to directly estimating software failure probability
    - This idea is limited by lack of data on previous software projects (the data are in general proprietary) and the applicability of this data.

Issues and Limitations of BBN Method (2)

- Complexity of BBN structure:
  - A BBN model developed using the structure of a standard on software development, e.g., BTP-14, tends to be very complex, with many node probability tables having to be estimated.
    - Detailed models require a detailed assessment, though there will potentially be little data available.
    - Alternatively, it may be more practical to develop a higher level model, though further exploration would be needed to demonstrate how to establish the relationship between the simplified model (e.g., an extension of the failure likelihood index method of human reliability analysis) and the associated guidance, and to determine the feasibility of quantifying such a model.

Issues and Limitations of BBN Method (3)

- Causal relationship representation in BBN structure:
  - A BBN model can be built based on the causal relationship of the relevant events in a specific application; the directions of the arrows represent the influence of the parent nodes on the child nodes.
    - If a single child node has multiple parent nodes, a high dimensional NPT or conditional probability density function must be formulated, which brings extra difficulties.
    - Such a structure is frequently modified by reversing the direction of influence, i.e., a single parent node with multiple child nodes.
    - These two types of modeling will be compared using simple examples to further evaluate their differences.

Issues and Limitations of BBN Method (4)

- Other Potential Limitations:
  - Unexpected or controversial results may surface in applying the collected software evidences to a BBN model. That is, the parameters and structure of a BBN may require revisiting and possibly revising if counter-intuitive findings emerge.
  - Hard evidence represents a realization of a particular state of a node, and soft evidence represents the discrete probability distribution of a node. Additional research is needed to understand the meaningfulness and value of using soft evidence.

Treatment of Uncertainty

- Parameter uncertainties are intrinsic to the distributions representing the nodes and node-probability tables
  - Assessing the distributions requires the availability of data or of knowledgeable experts, exemplified by the issue of converting qualitative information on software-development activities into software failure rates/probabilities.
- BBN modeling uncertainty can be addressed in a typical way, i.e., by determining the sensitivity of the PRA results to different assumptions or models.
  - Alternative assumptions and modeling approaches described in the previous slides can be evaluated (e.g., to consider the effect of the complexity of BBN models on the results, two BBN models can be developed, a simpler one and a more complex one, or alternative methods can be developed to compare to the BBN model [e.g., using the failure likelihood method developed for human reliability analysis]).

Procedure for Developing a BBN Model of an Example System (1)

- System familiarization
  - The needed documentation includes information on system design and operation, test- and operational-data, and a description of software development activities, such as debugging records and validation and verification reports.
  - Experts should be identified who are familiar with the software development of the example system. It is preferable to use a system that was developed following NRC guidance, that is, BTP-14, so that the model is more applicable to systems at U.S. nuclear power plants.
  - The major issue associated with this step is the potential unavailability of detailed information/documents of the example system, and of its software and associated development.
- Resolution of high-level issues
  - The most important issue to resolve is how to convert qualitative information on software development work into the software's probability of failure-on-demand.
  - The state-of-the-art in this area is very weak mainly due to lack of operational experience of similar software development projects.
  - There is a need for additional research, such as a benchmarking study of BBN applications to software reliability or data collection.

Procedure for Developing a BBN Model of an Example System (2)

- Development of high-level BBN model
  - The high-level structure depends on the resolution of the issue of converting qualitative information to software failure-on-demand, i.e., a direct or indirect conversion.
- Development of expert-elicitation approaches
  - Two types of application expert elicitations should be used. The first type employs experts (Type I experts) who are familiar with software-development activities and associated guidance and requirements to develop the generic BBN model
  - The second type of experts (Type II experts) are familiar with the development of the specific system under evaluation and provide system-specific answers associated with the observable nodes of the BBN model.
  - It is questionable whether adequate Type I experts can be identified. Type II experts should be more readily available, since the software developers should qualify as this type of expert.
  - Investigations are needed on how to best perform the expert elicitations.

Procedure for Developing a BBN Model of an Example System (3)

- Development of the complete BBN model
  - The low-level BBN model should account for the requirements specified in the guidance for developing software, that is, BTP-14 for U.S. NPPs.
  - It is desirable to keep the overall BBN structure as simple as possible because, even if a complicated BBN is built that correctly reflects the causal relationship between different nodes, not much can be gained from the model without sufficient supporting data.
- Expert elicitation to quantify node-probability tables
  - PRA experts should undertake this part of the study by eliciting Type I experts to quantify the generic part of the BBN model formulated in the preceding step.
- Collection of evidence on observable nodes
  - This step assesses the quality of software development activities of the specific software system under evaluation by collecting evidence associated with the observable nodes of the BBN model using available documentation on the system (e.g., debugging data).
  - It also encompasses eliciting Type II experts for answers to questions about the observable nodes of the BBN model.

Procedure for Developing a BBN Model of an Example System (4)

- Performing BBN analyses and generating results
  - Standard commercially available software tools can be used in developing the BBN model, and performing the needed analyses, such as generating a probability distribution for the probability of failure-on-demand.
- Performing sensitivity calculations to address modeling assumptions and issues
- Performing benchmark studies to evaluate the BBN model
  - The statistical testing results of the example system can determine if the prediction of the BBN model is consistent with the test results.

Current Status of the Example System - LOCS of ATR

- The Advanced Test Reactor (ATR) Loop Distributed Control System (DCS) software is a commercial software developed and provided by Metso Automation Max Controls, Inc. in compliance with ASME NQA-1 (Nuclear Quality Assurance).
- Related information acquired that can be used to support BBN model development:
  - The software quality assurance plan of the Loop Operation Control System (LOCS) upgrade and the review activities of the software development cycle.
  - Document illustrating the software V&V plan of the most recent LOCS upgrade project .
  - A Configuration Management Plan that contains some information about the software development activities.
  - Some details of a previous upgrade project of the LOCS and definition of the phases in the software development cycle.
- Additional information is needed to support the BBN model development (e.g., the actual V&V report of the most recent LOCS upgrade).

References

- [Delic 1997] Delic, K.A., Mazzanti, F., and Strigini, L., “Formalizing Engineering Judgment on Software Dependability via Belief Networks,” Sixth IFIP International Working Conference on Dependable Computing for Critical Applications, “Can We Rely on Computers?” Garmisch-Partenkirchen, Germany, IEEE Computer Society Press, pp. 291-305, 1997.
- [Eom 2009] Eom, H-S., et.al., “Reliability Assessment of a Safety-Critical Software by Using Generalized Bayesian Nets,” ,” *6th ANS Topical Meeting on Nuclear Plant Instrumentation, Controls and Human-Machine Interfaces Technologies (NPIC&HMIT 2009)*, Knoxville, Tennessee, April 5-9, 2009.
- [Gran 2000] Gran, B.A., and Dahll, G., “Estimating Dependability of Programmable Systems Using Bayesian Belief Nets,” OECD Halden Reactor Project, HWR-627, May 2000.

U.S.NRC
UNITED STATES NUCLEAR REGULATORY COMMISSION
Protecting People and the Environment

Development of a Context-Based Statistical Testing Method

Advisory Committee on Reactor Safeguards
Digital Instrumentation and Control Subcommittee

June 7, 2011

Tsong-Lun Chu
Brookhaven National Laboratory
(631-344-2389, chu@bnl.gov)

Brookhaven National Laboratory
U.S. Department of Energy

Outline of Presentation

- Background
- Contexts defined by a PRA
- A risk-informed strategy for performing statistical testing
- Demonstration of the feasibility of the risk-informed statistical testing
- Application to an example system

Background

- In 1997, a National Research Council committee completed a study requested by the NRC on application of digital instrumentation and control (I&C) technology to commercial nuclear power plant operations. It concluded that:

  1) "*Explicitly including software failures in a PRA for a nuclear power plant is preferable to the alternative of ignoring software failures*"

  2) "*As in other PRA computations, bounded estimates for software failure probabilities can be obtained by processes that include valid random testing and expert judgment.*"[1]

- In 2004, the ACRS Digital I&C Subcommittee pointed out the "system-centric" view of software failure and emphasized the importance of considering the contexts in which software failure can occur.

- In a recent NASA sponsored study, a context-based software risk model (CSRM) was developed and the software quantification method is based on statistical testing.

[1]Committee member Nancy Leveson did not concur with this conclusion.

Software Tests Performed for Licensing Purposes

- Software tests to support licensing are performed at different stages of development, and include unit tests, integration tests, system tests, factory acceptance tests, and site acceptance tests.
- The objectives are to detect bugs so that they can be removed, and to demonstrate the software can perform its intended functions.
- There may be requirements on the “coverage” of the tests, for example, every statement of the software has to be exercised.
- The test results in general cannot be used in estimating software reliability mainly because the tests were not sampled randomly from the operational profile of the software.
- To assess software failure probability (as opposed to finding bugs), statistical testing should be performed after, or as a part of, factory acceptance testing.

Statistical Testing of Safety-Critical Software in the Nuclear Industry

- Theoretical research has been performed by Miller [1992], May [1995], Littlewood [1997] and Kuball [2004].
- The tests performed for the Sizewell B reactor [May 1995] appear to be the only large scale “statistical” tests that have been performed.
- Research on white-box test methods were proposed by Zhang [2004] and Yang [2010].
- Costs associated with statistical testing may be significant.

Literature Review of Statistical Testing

- Black-box testing is done by sampling from the operational profile of the software and is a more straight forward method for quantifying software failure probability. It uses standard statistical methods for quantification. For example, using a uniform prior distribution, the posterior mean failure probability with no failures in n tests would be $1/(2+n)$.
- White-box testing requires, in addition, that tests be done in such a way that "coverage" of the paths and nodes is tracked and used in the quantification. That is, the frequency that a node or path is visited is used as the weight in calculating a weighted-average software failure probability.

Black-Box-Based Statistical Testing

- The operational profile needs to be defined such that random samples can be taken from it.
- Some research provides guidance on how to "randomize" when generating test cases (e.g., Littlewood and May).
  - The term "trajectory" (in the space of physical parameters) was introduced by May to represent the unique history of the plant physical parameters over time.
  - Each trajectory can be generated using a TH model of the plant simulating the occurrence of a plant condition in which a safety-critical system (e.g., RPS or ESFAS) is needed.

White-Box-Based Statistical Testing

- White-box testing takes into consideration internal structural elements (nodes and paths) of software and attempts to ensure all elements are covered.
- White-box-based statistical testing provides more detailed probabilistic information for the software; however, additional resources are needed as compared with black-box statistical testing.
  - May [1995] derived an expression for the mean failure probability $1/\Sigma(n_i \alpha_i)$ in terms of the number of tests (n_i) and "partitioning factor" (α_i) of the partitions of the input space. Based on this expression, the same number of tests has to be performed on each partition as is needed in black-box testing of the entire software, in order to demonstrate that the mean failure probability is of a certain value.
  - A similar observation can be made in terms of calculating an average failure probability in terms of the failure probabilities of individual paths or nodes.
  - Zhang [2004] found that it was not possible to test all paths of a Signal Validation Algorithm implemented in C language.
- The proposed risk-informed test strategy reflects some aspects of white-box testing, because it effectively partitions the input space. In addition, it takes into consideration PRA information that may reduce the number of tests needed to demonstrate software contribution to plant risks is "low enough."

Test Configuration of Statistical Testing (1)

- Test configuration needs to be as realistic as possible.
- Less realistic test configurations may not identify some possible software failures. For example, testing software without running it on the actual hardware may not be able to capture potential failures associated with interactions between hardware and software. In general, different test configurations represent different degrees of realism, and selection of the most appropriate test configurations is an area that would benefit from additional research.
- A practical limitation of statistical testing is the availability of realistic test configurations.

Test Configuration of Statistical Testing (2)

- Ideally, operating experience would provide evidence on software reliability. In reality, in situ statistical testing is not possible.
- In Oconee digital upgrade, the NRC did not consider the simulation-based testing tool SIVAT an acceptable validation test tool. The ERBUS test machine that tests the actual hardware and software was used instead.
- On the other hand, it is not practical for the NRC to develop its own tool for testing software. Vendor developed tools have to be used.

Contexts Defined in a Probabilistic Risk Assessment

- A PRA models the systems that are needed in mitigating accidents. The contexts/conditions in which a protection system is needed are clearly identified in the PRA.
- Recognizing that software failure is very context dependent, it is natural to try to estimate a failure probability for each of the contexts/conditions.
- Due to expected high reliability of protection system software, it may not be possible to perform sufficient statistical testing for each such condition to estimate the true software failure probability.
- A more practical approach may be to perform enough tests to demonstrate that the contribution of software failure is “low enough,” taking into consideration that, in many failure scenarios, other failures (in addition to the software failure) have to occur in order to result in core damage.

A Risk-Informed Testing Strategy

- Perform tests on a protection system (e.g., RPS or ESFAS) for conditions/scenarios defined in the plant's PRA to demonstrate that the software's contribution to plant risk is small enough.
  - Other failures in a PRA sequence/cutset may help reduce the number of tests without failure needed to demonstrate an acceptable software failure probability for a given context. For example, for those plants with a small contribution to risk from anticipated transients without scram, the number of tests needed for the RPS software may not be very large.
  - It may be easier to show the contribution of software failure is small comparing to the contributions of other contributors, for example, demonstrating that the probability of loss of a safety injection signal due to software failure is small compared to the probability of failure of the safety injection system due to mechanical failures.

Generation of Test Cases

- Use the contexts defined by the plant PRA. For example, each initiating event defines a context for the RPS, and each event tree sequence that includes an ESFAS function defines a context for the ESFAS.
- For each context, generate multiple test cases (trajectories) by considering variation in the initial condition of the plant, variation in the initiating event (e.g. LOCA location and size), timing of preceding failures in the sequence, if any, and sensor inaccuracies and delays. The required number of test cases without failure depends on the software reliability to be demonstrated.
- Input the trajectories to the system being tested and determine if the needed protection signal is generated.

Demonstration of Feasibility of Risk-Informed Testing Strategy

- Determine a generally acceptable increase in CDF for the plant (e.g., based on R.G. 1.174).
- Identify all core damage sequences in which a protection function is failed due to loss of an actuation signal.
- Determine the upper bound on the mean software failure probability such that when added to the digital hardware failure probability, the CDF of each sequence remains at or below the CDF for the corresponding sequence involving the analog protection system
  - Assuming the digital hardware failure probability remains the same for all such sequences, then the upper bound on the mean software failure probability will also remain the same for all sequences.
- Assume the acceptable increase in CDF is due entirely to these sequences, and optimize the allocation of the acceptable increase to these sequences in a manner that minimizes the total number of tests needed without failure to demonstrate the mean software failure probability is low enough in all cases so that the cumulative plant risk does not exceed the generally acceptable increase.
  - The premise is that in sequences of lower frequency, the software failure probability would not have to be demonstrated as low as in higher frequency sequences.

Key Assumptions of Statistical Testing Method

- Operational profile definition and characterization can significantly affect the results.
- Realism of test configuration can affect usefulness of the test results.
- “Memory” of digital systems and internal states can affect the tests’ ability to capture certain failures.
- Internal component failures are not accounted for.
- A very large number of tests may potentially be needed.

Improvements in Addressing the Desirable Characteristics

- The proposed risk-informed approach to the black-box statistical testing method may better address the following desirable characteristics:
  - *Characteristic 3: The method allows for consideration of the specific operating conditions of the software.*

    The proposed risk-informed testing strategy of this study does take into consideration specific contexts/accident scenarios.
  - *Characteristic 7: The method has been verified and validated.*

    The successful application of the proposed statistical testing approach of this study to an example system would help demonstrate the method.
  - *Characteristic 8: The method is capable of demonstrating the high reliability of a safety-critical system.*

    The risk-informed testing strategy proposed in this study might show that high reliability may not be necessary in demonstrating that the software's contribution to the plant's core damage frequency is small enough. A feasibility study is recommended that estimates the total number of tests that must be performed to determine the practicality of the approach.

Application to An Example System

- Loop Operating Control System (LOCS) of the Advanced Test Reactor (ATR) at Idaho National Laboratory is the example system.
- The ATR has five in-pile tubes that are equipped with pressurized water cooling loops.
- The LOCS controls the loop cooling systems of the in-pile tubes, and generates a reactor trip signal upon detecting abnormal conditions of the loop cooling systems.

Statistical Testing of LOCS of the ATR

- Demonstrate how statistical tests can be performed to estimate software failure probability.
- Perform TH calculations of scenarios associated with the in-pile tubes to generate test cases that require a reactor trip and save the results in a file.
- Provide the file containing the results to a test configuration (possibly the simulator) that reads the test cases and records the outcome of the tests in terms of successes and failures.
- If sufficient PRA information is available, explore risk-informed statistical testing approach for this system.

United States Nuclear Regulatory Commission

Protecting People and the Environment

Summary and Future Interactions

Advisory Committee on Reactor Safeguards
Digital Instrumentation and Control Subcommittee

June 7, 2011

Alan Kuritzky
Division of Risk Analysis
Office of Nuclear Regulatory Research
(301-251-7587, Alan.Kuritzky@nrc.gov)

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Summary (1)

- Current focus
  - Incorporation of software failures into digital instrumentation and control (DI&C) reliability models for use in nuclear plant probabilistic risk assessments (PRAs)
- Candidate method selection
  - Review of common quantitative software reliability methods (QSRMs) and comparison against a set of desirable characteristics indicated that no method meets all of the characteristics nor clearly stands out as the most appropriate.
  - Based on insights from the review and comparison, and more detailed assessments of the software reliability growth method, Bayesian Belief Network (BBN) method, and test-based methods, the black-box statistical testing method was identified as the preferred approach.
  - Due to limitations of test-based methods and to account for the quality of software lifecycle activities, a prior distribution will be developed using the BBN method, and it will be updated by Bayesian inference using the statistical testing results.

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Summary (2)

- Additional considerations
  - Since developing and quantifying a detailed BBN model can be very resource intensive, and the evidence needed for quantification may be lacking, both a simplified and detailed BBN model will be developed.
  - The statistical testing results will also be used to perform a Bayesian update of a non-informative prior, as an additional standard for comparison.
  - In conjunction with the black-box testing method, a risk-informed testing profile will be explored to help account for the software context and to minimize the testing burden.
  - The case studies will be performed in collaboration with the Korea Atomic Energy Research Institute (KAERI).

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Future Interactions

- Draft NUREG/CR on application of BBN method to example system released for public comment (tentatively, August 2012)
- Internal review, external review, and public comments to be incorporated into draft final NUREG/CR on application of BBN method to example system (tentatively, December 2012)

Brief ACRS DI&C Subcommittee on draft NUREG/CR in Sept/Oct 2012 timeframe? (no letter requested)

- Draft NUREG/CR on application of black-box statistical testing method to example system released for public comment (tentatively, February 2013)
- Internal review, external review, and public comments to be incorporated into draft final NUREG/CR on application of black-box statistical testing method to example system (tentatively, July 2013)

Brief ACRS DI&C Subcommittee on draft NUREG/CR in Apr/May 2013 timeframe? (no letter requested)