# MEMORANDUM

**TO:** Parviz Adib

**FROM:** David Patton
Pallas LeeVanSchaick

**DATE:** May 16, 2006

**RE:** Real-Time Co-optimization Proposal

This memo provides our comments on the draft WMS whitepaper regarding co-optimization of energy and reserves. First and foremost, the proposal in the WMS whitepaper is not a proposal to co-optimize the energy and ancillary services markets. Rather, it is a *sequential* optimization, with reserves selected on an hourly basis ahead of the operating hour and energy dispatched on a five minute basis. Real-time co-optimization requires that energy and reserves be selected and priced on a five-minute basis to accurately reflect system conditions.

The sequential optimization would increase the efficiency of the markets in comparison to the current nodal protocols, which do not provide for any optimization between energy and reserves during the operating day. However, true co-optimization would increase market efficiency of Texas Nodal markets substantially more than would sequential optimization. The NYISO operated a system substantially similar to this proposal, reallocating operating reserves on an hourly basis but dispatching the system on a five-minute basis. In February of 2005, the NYISO implemented co-optimization on a five-minute basis and it has resulted in substantial benefits in comparison to the sequential optimization.

As discussed below, in comparison to the sequential optimization process that is proposed, co-optimization would:

- Improve the allocation of resources between reserves and energy.
- Reduce the amount of make-whole uplift costs.
- Facilitate efficient pricing during shortage conditions (i.e., “scarcity pricing”).

Additionally, we believe that the resources needed to implement the sequential optimization process proposed in the whitepaper likely exceed those needed to implement co-optimization for the following reasons:

- Most vendors' real-time market software already include various forms of reserve constraints, so limited additional resources would likely be needed to include real-time co-optimization in the market model.

- Alternatively, sequential optimization requires the implementation of a third model (day-ahead, hour-ahead, and real-time). Although the hour-ahead model would be very similar to the real-time model, it would still need to be developed and tested. Additional data interfaces for other systems would be need to be developed and tested that provide input to and accept output from the hour-ahead model.

- Both approaches require the development of billing and settlements processes and interfaces to implement a multi-settlement system for ancillary services (day-ahead and real-time settlement). Once the resources are expended to develop the multi-settlement functionality for ancillary services, the difference in costs between settling real-time ancillary services on a 15-minute basis (co-optimization) versus an hourly basis (sequential optimization) would likely not be substantially different.

Based on these observations, we recommend that ERCOT and vendors compare the feasibility, associated costs, and potential timeline of implementing sequential optimization and real-time co-optimization. Depending on the results of this comparison, we recommend the adoption of one of the following alternative approaches:

1. Implement real-time co-optimization as part of the phase 1 implementation of the nodal markets. As discussed above, the primary work would be in the billing and settlement systems. These systems will already be designed to support 5-minute dispatch and 15-minute settlements in real-time. Co-optimization should simply increase the quantity of 5-minute and 15-minute data that must be processed, but should not otherwise significantly change the requirements for the billing and settlements.

   However, if ERCOT concludes that phase 1 implementation of real-time co-optimization is not feasible, we recommend the following alternative:

2. Implement co-optimization the year following the phase 1 implementation of the nodal markets. This approach would help ensure that the phase 1 implementation is not delayed and would capture the full efficiency benefits of co-optimization shortly thereafter. This is superior to implementing the sequential optimization because the resources needed to implement the sequential optimization can be utilized to develop the systems and software to do co-optimization prior to the phase 1

implementation. This would ensure that co-optimization would be ready for implementation shortly after the phase 1 implementation.

**Efficiency Concerns Regarding Sequential Optimization**

The sequential optimization process proposed raises efficiency concerns when compared to real-time co-optimization. These concerns generally arise because the hour-ahead model cannot accurately predict conditions in real-time and would, therefore, make mistakes in its scheduling of resources that are inefficient. These efficiency concerns cannot be fully addressed in any sequential system; they can only be minimized by seeking to make the hourly model as accurate as possible. However, it is impossible for the hourly model (regardless of how accurate it is) to fully capture the rapidly changing conditions that can occur on a five-minute basis, particularly during the high-ramp hours that occur in the morning and evening.

As a result, unless a third settlement is performed (day-ahead, hour-ahead, and real-time) in which the energy and AS prices are financially binding, an hour-ahead model that reallocates reserves would result in substantially larger uplift costs associated with make-whole payments to generators compared to five-minute co-optimizaton. The make-whole uplift, which would occur under either a sequential or real-time co-optimization, is designed to ensure that day-ahead AS suppliers that are de-scheduled in the hourly process or real-time AS suppliers that are scheduled in the hourly process are held harmless relative to selling energy at real-time prices. A third settlement would be much more costly than settling the AS markets on a 15-minute basis.

**Scarcity Pricing and Sequential Optimization**

An earlier draft of the whitepaper included an assertion that real-time co-optimization hinders scarcity pricing. This misconception may account for why the 5-minute co-optimization of energy and reserves is not seen by some market participants as significantly superior to the hourly sequential optimization proposed in the whitepaper.

One of the most significant benefits of co-optimization is that it sets prices that reflect the trade-off between energy and reserves during shortage conditions, both system-wide and regional shortages. In order for the market to solve under conditions where resources are insufficient to meet both the energy and operating reserve requirements, the model must include a “penalty factor” that allows it to violate the reserve requirement. If this penalty factor is set appropriately to reflect the value of the reserve to the system, it will generate efficient prices during shortage conditions (i.e., it will be included in the prices for both energy and reserve when the system is in shortage). Although the use of penalty factors is not currently included in the Nodal Protocols, a real-time co-optimized model will not solve without them when the system is in shortage because there will be no feasible solution (i.e., no way to satisfy both the operating reserve and energy requirements).

With regard to the assertion that co-optimization may hinder scarcity pricing signals, this can only occur in cases where there is not authentic system-wide or regional scarcity. In these cases, a competitive market should fully utilize available resources in the most economic manner possible rather than allocating resources in a sub-optimal manner and creating the illusion of a shortage.

**Scope of the Market**

We agree with the whitepaper that it is appropriate to co-optimize energy with reserves rather than reserves and regulation. As pointed out in the whitepaper, frequent reallocation presents more operational issues for regulation than for reserves. Secondly, there is much more flexibility to reallocate capacity between reserves and energy than there is between regulation and the other two. This is partly due to the specialized nature and costs of providing regulation as opposed to providing reserves. The costs of providing reserves are primarily the opportunity costs of not providing energy, therefore making the energy dispatch an essential consideration. Furthermore, since the quantity of reserves needed is greater than the quantity of regulation needed, the demand for regulation has less impact on the cost of satisfying energy demand than does the demand for reserves.

One potential benefit of co-optimization is that it can help improve the accuracy of price signals during non-spinning reserve deployments. In the current nodal protocols, ERCOT would deploy non-spinning reserves manually when it is in danger of running short of energy. During these periods of shortage, the additional supply that is added to the energy market has the effect of dampening scarcity price signals. However, in a market where energy and reserves are co-optimized, the deployment of reserves for energy needs is based on the economic trade-offs between maintaining reserves and satisfying demand. Thus, we recommend that non-spinning reserves be included in the co-optimization and that the deployment be based on clear economic criteria under the co-optimized dispatch rather than the discretion of the operator.