

May 16, 2006

Review of Efficient Treatment of Ramping in the Context of the Ontario Market Prepared for the IESO



NERA

Economic Consulting

1166 Avenue of the Americas
New York, New York 10036
Tel: +1 212 345 3000
Fax: +1 212 345 4650
www.nera.com



MMC Marsh & McLennan Companies

Since the Ontario market went live on May 1, 2002, the pricing algorithm which has compensated generators and charged loads has, as a temporary fix for a perceived problem, treated all units as if they could ramp 12 times faster than they report. We have been asked by the IESO to address the question of whether that temporary fix should remain, and, if it should be replaced, by what¹. To that end, we have prepared this report, which discusses ramping questions in general, discusses six possible approaches to considering ramping in the prices in the IESO administered Ontario spot market, and evaluates those approaches against a relevant set of evaluation criteria. We will begin with a description of the ramping issue and a description of the *status quo* solution. Section II provides the proposed alternatives with a brief discussion of the rationale for each. Section III provides a list of six criteria by which the market rule will be judged, and Section IV evaluates the *status quo* and alternatives against the criteria.

Our main conclusions are:

- First, the assertion that the ramping is currently uncompensated reflects a misunderstanding of the Ontario market. Any generator who would prefer to operate at a set MW level rather than ramp is free to do so by bidding a low price for energy below that point and a high price for energy above that point. When the IESO chooses to ramp that unit, that is either reduce the loading in a period so the unit can provide ramping in a subsequent period or dispatch the generator above its preferred set point, the generator earns a potentially large payment for having its preferred position altered -- the difference between its bid and the market clearing price. Similarly, a unit which would prefer to remain offline rather than be brought on-line to ramp can be compensated by bidding a high energy price.
- Second, energy is paid based on an optimization algorithm. The closer the optimization algorithm is to the way the system is dispatched, the more efficient the price signals will be. This conclusion must be modified to reflect the fact that the decision not to pay locational prices has already skewed the pricing dispatch from an operational dispatch, so it is not *necessarily* true that all improvements in the pricing algorithm bring about improvements in efficiency, but it is probably true in Ontario. The conclusion to be drawn from this is that Multi-Interval Optimization (MIO) is probably to be preferred over myopic optimization and 1X ramping is probably to be preferred to 12X ramping
- Third, the current value of ramping is low. The evidence for this is that ramping can currently be accommodated without a large amount of constrained-on payments. Given the probable low current value of ramping, direct payments for ramping are probably not warranted at the present time. The optimal level of such payments would probably be quite small. That said, failure to pay for ramping directly (as opposed to the current method which pays through constrained-on payments) has efficiency losses through the constrained-off payments which the constrained-on payments entail. Should ramping become valuable in the future, a problem could develop which might suggest setting up a structure which makes direct payments today.

¹ We note here that this issue has been combined with many others, such as whether or not the current dispatch algorithm issues an excessive number of spurious instructions. We are not addressing that question here directly.

- Fourth, if such payments are to be made, the level should be determined by the incremental costs of ramping rather than by an arbitrary calculation.
- Overall, the question of what should be done is bound up with the nature of the priorities in the Ontario market. In our opinion, the lack of a day-ahead market for scheduling and unit commitment and the failure of the market to incorporate important transmission constraints into pricing represent first-order problems, while pricing issues around ramping issues are of a much lower level of current concern. This does not mean that they should not be addressed, of course, but modifying the pricing algorithm to account for ramp will not have dramatic effects in the near-term in Ontario. That situation may change, however, with the retirement of the coal units and the substantial expansion of nondispatchable generation in Ontario.

I. The Ramping Question

A. What Is Ramping?

Interconnected generators cannot instantaneously change their output levels, either up or down². If system load changes by more than the given generators on the system can change their output, the system is said to be *ramp-constrained*. Since at any time, the quantity of electricity consumed must equal the quantity generated,³ system stability requires that any such conditions be highly transient. In the limit, load will have to be shed to keep the system stable.

In the short-term operation of the Ontario system, the system is rarely ramp-constrained because the IESO (using a multi-interval dispatch algorithm) plans ahead to ensure that an adequate amount of ramping capability is on the system at any given moment. For the most part, these actions occur at regularly recurring times; in particular, as households and businesses begin to consume electricity every weekday morning, large quantities of load are added fairly quickly. On average in 2005, Ontario load rose almost 1700 MW on weekdays between 6 am and 7 am and about the same amount between 7 am and 8 am, with another 1000 MW between 8 am and 9 am. This effect is even more pronounced on some days: on 06/27/05, demand rose by 5570 MW in this three hour period, exceeding the normal weekday ramp by over 1000 MW⁴.

Units vary greatly in their ramping capacity. Hydro units, for example, are generally able to ramp quite quickly, while steam units are often quite slow. Modern combined cycle gas turbines are often not built to ramp efficiently. This difference in ramping capability between units means that the economic dispatch can depend on not just the energy and reserve offer curves, but also on ramping capability. The necessity to meet a morning's demand growth may cause some combination of the following actions:

² Of course, if transmission pathways fail, any generator can drop to zero system input.

³ This includes losses and assumes that voltage and frequency are held steady. In actual operations, as loads are added to the system, frequency falls until the new generation rebalances the system.

⁴ While the morning ramp up is a prominent cause of ramp, it is not the only one. Hourly changes in interchange quantities also cause ramp. Ontario, like all other systems, limits the interchange amount at the hour to accommodate the internal ramping capabilities.

1. Units which are capable of fast ramps but have low energy bids may be dispatched below their full output even when energy prices are above their bid. Thus, they will be constrained off and, all else equal, receive a CMSC payment equal to the difference between the energy price and their bid, multiplied by the quantity of capacity held back.
2. Units which are capable of fast ramps but have high energy bids may be dispatched at minimum load even when energy prices are below their bid. Thus, they will be constrained on and, all things equal, receive a CMSC payment of the difference between their bid and the energy price, multiplied by the minimum load quantity.
3. Units which are incapable of ramping quickly may be ordered to take out-of-merit actions as well. A unit may be asked to ramp down, for example, well before prices have turned down in the evening, since otherwise they will be on at too high a level when prices fall well below the unit's costs.
4. Units which are incapable of ramping quickly may be asked to gradually ramp up out-of-merit so as to meet the system peak when it finally comes.

Up to this point, it is unclear that there is any problem at all. Generators who are dispatched out of their unconstrained energy merit are made whole through CMSC payments, so they are paid at least the published price vector for energy and reserves. The problem, if any, comes from the energy price itself. To what extent should energy prices reflect the ramping constraints?

B. Price Formation In The Ontario Market

The Ontario energy price is the *shadow price* of an optimization program. A shadow price is simply the marginal change in total cost with respect to the marginal relaxation of a constraint, holding the other constraints constant. One constraint is that total quantity of energy generated has to equal total quantity supplied. It is this energy constraint that is used to set the Ontario energy price. This energy constraint is almost always binding⁵. Thus, we imagine that in a particular five minute interval that load had been 1 MW higher. We then imagine redispatching the system holding all other constraints constant and ask how much more costly it is to supply that extra MW for that five minute period.

Often, the marginal cost of energy is just the cost of some generator, not currently fully loaded, being instructed to slightly increase his output. In that case, which is the easiest to think about, the marginal cost is the marginal offer price of the marginal generator. We often speak as if that is the only case, but it isn't. One other case which has been discussed in the past results from the co-optimization of energy and reserves. When energy and reserve are co-optimized, it is possible that an increase of 1 MW of load will cause a price increase of more than the marginal offer of the cheapest generator not fully loaded, because the unused MWs of that generator were already being used to supply reserves. Thus, we must either use a MW of energy from another, more expensive generator, or use the original generator for energy but also procure 1 MW of reserves from the least expensive offerer of reserves with reserve capacity available. In this

⁵ An exception would be if generation could not ramp down as quickly as load was falling.

case, the marginal cost of energy includes all the other actions related to all other constraints on the system needed to keep those constraints within bounds.

Just as the co-optimization of energy and reserves affects the price of energy in ways that are not obvious at first glance, the enforcement of a ramping constraint affects the price of energy. Complicating the situation somewhat, another set of constraints is also quite important in this calculation: minimum load constraints.

Imagine an increment in load in a particular interval. Optimal dispatch to this increment of load may consist of some or all of the following:

- Turning up the least expensive generator on the system with additional capacity to sell
- Bringing on a new unit and dispatching it at minimum load
- Turning off or ramping down a unit to accommodate the minimum load of the new unit
- Reallocating reserves among units capable of providing reserves
- Altering imports at the start of the hour and backing off other Ontario-based units across the hour

Note that the last action is only feasible on a look-ahead basis, *i.e.* if we knew at the time that the interchange schedules were set that we were going to need another MW in this interval. This distinction will be important later.

The basic point is that we need two things to calculate an energy price: a state of the system at some time and an optimization algorithm which redispaches the system assuming a small change in some output, holding the others fixed, and reports the cost of meeting that change. Thus, before we can answer the question of whether and how energy prices ought to reflect ramp constraints, we need to explore these two components in slightly more detail. First, we will look at what sort of constraints are modeled and second we will look at the definition of the initial state of the system.

C. Modeling Constraints and Their Impact on Price Formation

A natural first supposition is that if a constraint exists in the real world, the constraint should be in the model. *First*, we should note that this is not really possible. All models of electric dispatch leave out some real-world constraints and in fact impose others which don't really exist. This is because the true optimization under every known constraint is just too hard a mathematical problem. A real AC power-flow model of the system would take days to optimize, so we know that we are bound to leave off some constraints in calculating price just to get an answer in a reasonable period of time.

Second, however, we are free, if we wish, to ignore constraints we know to be valid and that we really can solve in reasonable time. An obvious example of this in Ontario is transmission

constraints. Energy prices in Ontario are calculated in a so-called unconstrained dispatch, by which we mean unconstrained by transmission constraints. In an explicit counterfactual, we assume that an increment of power could be produced by any Ontario-based resource, even though we know that some resources would not be able to achieve demand-supply balance because of transmission constraints.

It is the decision to ignore constraints in the pricing algorithm which primarily causes CMSC payments. This flows from the fact that the scheduling and dispatch optimization process optimizes respecting a number of constraints, including transmission constraints, for which shadow prices are ignored in the market pricing rules. If the shadow prices were applied for all constraints, then we would tend not to make CMSC payments because every unit would receive market prices fully consistent with its dispatch instructions: the prices received would always either be greater than or exactly equal to the dispatch costs offered.

The existence of constraint payments create two inappropriate signals. *First*, they put a wedge between the observed energy price and the total price of power consumers pay. This makes it difficult, if not impossible, to send loads and generators the correct signal. *Second*, however, and probably more important, *whenever we make a constrained-off payment, we pay a generator whose output was in fact not needed*. Note that whenever we constrain 1 MW on, we must constrain some other MW off somewhere in the system. For example, when we ignore locational differences in pricing, we may make CMSC payments to those located on the low side of the constraint. Likewise, when we ignore ramping constraints, we may make payments, either directly or indirectly, to those units whose inflexibility has given rise to the constraint which causes us to alter the dispatch. If all generating units were capable of instantaneous ramps there would be no ramping constraints and no need to deviate from an economically optimal dispatch.

Since constrained-on and –off payments arise from the use of a dispatch in pricing which is not the same as the dispatch used in actually running the system, aligning the pricing dispatch with the actual dispatch lowers uplift costs. Complicating this story in answering the ramping question is that we are already violating this principle in transmission congestion.

What about the generators who are constrained on? As we understand it, generators have argued that those who are called to ramp are not properly compensated for their services. This cannot be correct. No generator can be constrained on for a payment which does not compensate him, since he is free to modify his bid to ensure that if he is constrained on, he is content with the price received. It has been objected that this violates the principle that all generators bid at marginal cost. But there is no such principle in the Ontario system today, nor was there such a system at the beginnings of the system design.

It is true that bidding the cost of ramping into price will mean that some generators, at some times, will not be taken when they otherwise would have. But this is an efficient result, since the ramping which would have been performed by this unit is now performed by some other unit, which was willing to bear the combined costs of energy, ramping, and possibly reserves, at a lower price than the unit displaced. This is simply the principle of competition at work.

Although using constrained-on bids to compensate generators must be compensatory, it may well be inefficient, however, for several reasons:

- First, the process of bidding around constraints will inevitably be imprecise. The costs of ramping are probabilistic in nature, particularly those relating to wear-and-tear, so that bids cannot fully reflect costs. Units which might be optimally used to ramp will not be taken because their bids (not their costs) were slightly too high.
- Constrained-on prices give no opportunity to earn payments higher than one's bid; consequently, one might expect bids slightly higher than one's indifference point, raising costs to consumers.
- Constrained-on payments require generators to have a good knowledge they will be constrained-on. While the repeated nature of the ramping problem helps in this respect, generators with large portfolios can afford to invest in market models which more precisely predict how they should bid. Consequently, large entities are favored over small entities.
- Constrained-on pricing is discriminatory. It pays different amounts to two generators who ramped exactly the same.

In essence this paradigm reflects a pay-as-bid pricing scheme for ramping. There are good reasons to reject pay-as-bid pricing in the energy market⁶ and most of those reasons should follow over into the ramping problem.

D. The Effect of Starting Conditions

As we said above, the calculation of shadow prices requires a state of the system at which prices are to be evaluated. This is nothing more than the state in which the system is found. There are three possibilities of how the state of the system got to be that way. The first is that the initial state is itself the outcome of the same optimization routine that will be used to calculate the shadow prices. The second possibility is that the state of the system has been determined by an optimization model which is different than the one used to determine energy prices. The third possibility is that the state of the system is different, either because generators have failed to follow instructions or because system operators have explicitly overridden the optimization model.

While the mechanics of the calculation of shadow prices are the same however the initial state arose, the implications can be quite different in two respects. *First*, so long as the optimization is accurate, costs will, in aggregate, be lowest when the initial state was an optimum. This does not mean that costs could not be higher in some periods and lower in others along the optimal path – for example, the optimal ramping strategy may involve lower costs in the period leading up to the ramp as the facilities which will be used to ramp are loaded to their minimum loads, with higher prices later as those units are actually called on to ramp.

⁶ See <http://sparky.harvard.edu/hepg/Papers/kahn-cramton-porter-tabors-blue-ribbon-panel-report-to-calpx-01-01-23.pdf>

Second, using the optimum as the initial condition preserves transparency. To the extent that system operators' actions have overridden the optimization outputs, and energy prices are different in some interval (either higher or lower) there will always be the question of whether the system operators' actions were avoidable. The justification for such action must always be that the optimization algorithm's shortcuts had put system security at risk.

Beyond these concerns, the initial state, whatever it is, can have a dramatic influence on costs. To take a simple example, the failure of a large generator through forced outage will use up reserves at a much faster rate than was anticipated when the optimization program ran. It is then unsurprising that in the case of a generator trip, reserve prices might spike quite sharply when in the absence of the failure, reserve prices would have been quite low.

Similarly, ramping costs may be quite modest along the optimal path as only minor out-of-merit actions need to be taken in normal circumstances. Should a unit brought on for ramping purposes suddenly fail, however, the price of energy might rise quite sharply, indeed at the limit to the value of lost load, depending on the spare capacity available to provide the ramp. Just as generator failures lead to reserve shortages in normal periods, generator failures in periods of ramping are all the worse since the units which can ramp effectively form a small subset of all units.

The starting conditions must be relied upon in order to calculate an energy price, however the starting conditions, in and of themselves, can yield CMSC payments. This might be because of non-convexities in the optimization (e.g. we just started an expensive unit because we know we'll need it in three hours, or because actual operations for whatever reason didn't match instructions). Previous decisions made, regardless of how they were made, can impact the prices calculated going into any particular 5 minute period. An issue then, is the extent to which intertemporal aspects should be included in the pricing method, if at all.

E. The Status Quo: 12x Ramp, Myopic Optimization, No Minimum Load Constraints

Current pricing in the Ontario system uses an optimization model with three essential⁷ features:

- 12X Ramping: While the model includes ramping constraints, all units are presumed to be able to ramp 12 times faster than they actually can. The effect of this is to almost completely eliminate the ramping constraint.
- Myopic Dispatch: In optimizing the system, the algorithm does not look ahead to see where the system needs to be several intervals ahead. Consequently, it does very limited planning for future ramps and is essentially “taken by surprise” when ramping is needed.
- No minimum load constraints: All units are assumed to be capable of providing very small amounts of energy. Thus, while a unit is assumed to have ramping capability substantially in

⁷ None of these descriptions is a complete description of the dispatch algorithm used for pricing. None of them, for example, use transmission constraints. All of them co-optimize energy and reserves.

excess of its true capability (perhaps decreasing prices), it is not required to be dispatched to minimum load which displaces cheaper generation (and thus increases prices)..

The historic rationale for this pricing algorithm is piecemeal. Myopic dispatch was part of the original design of the Ontario system. If the system actually followed myopic dispatch while units bid their short-run marginal costs, prices would be very high, because the system would be constantly under prepared for well-understood eventualities like the need to bring units on out of energy-bid merit in order to ramp. The theory then suggested that to take advantage of these high prices, units would alter their own bids – fast-ramping units would lower their bids at exactly the time they needed to be loaded in order to take advantage of the higher prices. This in turn would lower those prices as adequate ramping capacity was on line when needed.

This vision was never implemented. In testing of the Ontario system, the dramatic rise in prices under myopic dispatch and marginal cost bidding dissuaded stakeholders from testing the theory that deviations from marginal cost bidding would arbitrage these high prices away. Instead, the myopic dispatch remained, but the 12X ramp rates were employed in order to “correct” the price rises. Once all units were free to ramp, the increase in prices from the morning ramp went away.

Finally, the use of this pricing algorithm has another feature. Because the ramping parameters are overstated in the algorithm, the system cannot be run with the implied dispatch of this algorithm. Instead, system operators routinely use a multi-interval algorithm which looks forward to use the units which actually can ramp rather than those which cannot. In addition, actual system operations obey the minimum loading constraints. Thus, the myopic optimization routine uses a set of initial conditions which are far from what look like (to it) to be an optimal state of the system.

Thus, we see that there will be units who will be paid constraint payments even though they cannot be run owing to their inflexibility in meeting the morning ramp. Just because this is true does not mean it needs to be changed. There are units who are paid constraint payments even though they cannot be run owing to transmission constraints, and there is no pending proposal to change this situation. To see whether change is warranted, and, if change is warranted what the change should be, is the subject of the remainder of the paper.

II. Alternative Proposals for Price Formation

While the *status quo* is still a possibility, we have been asked to consider the following five alternatives.

A. 1X Ramp Rate, Myopic Optimization, Minimum Load Constraint

The next proposal is, not surprisingly, to return to the original vision of the market. In this model ramp rates would be input, minimum load constraints respected, and myopic pricing dispatch employed. While, as noted in the previous section, this model predicts very high prices in the morning under marginal cost bidding, the philosophy which underlies this pricing model is that arbitrage possibilities will not last. Units which see these high prices will lower their bids to be dispatched earlier. This in turn will lower shadow prices during the morning ramp.

The effect of the minimum load constraint should be noted here. The effect of minimum load constraints is, in general, to lower prices. This is because the action of dispatching generators to minimum loads displaces higher cost generation which otherwise would have been brought online. So the net effect of adding both sets of constraints is not clear even when generators have bid at marginal cost.

B. 1X Ramp Rate, Multi-Interval Optimization, Minimum Load Constraint

If generators optimally skew their bids to arbitrage between intervals then the resulting dispatch will look just like one in which a central planner took all intertemporal issues into account. If one distrusts this notion, one can substitute an explicit multi-interval planning process (MIO dispatch). Thus, the optimization algorithm looks ahead far enough to plan the system. Unlike the myopic dispatch, so long as no contingencies develop, this algorithm is never taken by surprise by regularly recurring and largely predictable events like the morning ramp.

Another way of looking at this algorithm is to argue that it is what system planners actually do. If generators have not self-scheduled (through their own bidding) enough resources to provide for the morning ramp, the planners do it for them.

One attribute of a multi-interval optimization is that the actions taken to relieve a constraint in one interval have price effects in other intervals. For this reason, some have proposed assigning the payments to the periods which *caused* the problem. Thus, if under the MIO, a ramping requirement in the one interval in an hour causes prices to be higher several intervals before that, the increment in prices is assigned to the later period because it was the ramping which caused the increase in price.

This method, the so-called modified incremental method, while perhaps fairer in concept, would be highly inefficient to implement. In particular, all units are co-optimized for the provision of energy and reserves. If units are compensated more highly in later periods than is justified by the shadow prices, the balance between energy and reserves will be upset. Units called on to provide reserves in that period will see a lost opportunity as higher profits would have been earned had they been allowed to provide energy. These equilibrium conditions limit the ability to shift shadow prices intertemporally.

Another proposed method, the highest slice method, yields a price equal to the highest unit called on to run. This method is the same as the incremental method in many periods, but will yield higher prices in intervals where MIO is taking actions to plan for future demand patterns. The problem here is that it is possible to construct an example just before the ramp where energy is very inexpensive. In essence, if we could incentivize consumers to consume **more** energy shortly before the need for ramp, we would need less ramp. The highest slice method would charge high prices in this period, since the preparation for the ramp might have several units turned on out-of-merit. This could indeed exacerbate the ramping problem as consumers are dissuaded from using energy in a period in which we would like them to consume more, not less.

C. 12X Ramp Rate, Myopic Optimization, Payment to Ramping Units

An objection has arisen that the status quo fails to do anything more than leave rampers called on out-of-merit whole, *i.e.*, they are not paid to ramp. This argument presumes that they have bid marginal cost. In any case, there are two proposals on how to pay them:

- Take the difference between the 1X Myopic Price and the 12X Myopic Price, where positive, and pay this quantity to those units whose dispatch instruction changed over the interval. For the morning ramp, only those units called on to increase production during the ramp will be paid this price.
- Pay a fixed sum per MW ramp to all generators who vary their dispatch over an interval.

D. 12X Ramp Rate, Myopic Optimization, Ramping as Ancillary Service

Where constraints are economically important and potentially competitively supplied, another choice is to make a market for relieving the constraint. Ontario has created a market for relieving the energy constraint (as have all competitive markets) and a market to relieve the reserve constraint (as have many markets.) A logical extension, therefore is to create a market to relieve the ramping constraint. Just as generators get paid for energy and reserve, generators could also get paid a specific price for ramping in a given hour. The optimization software would co-optimize, producing a reserve price (being the shadow price on the reserve constraint), an energy price (the shadow price of the energy constraint) and a ramping price (the shadow price of the ramping constraint). Taking into account the total requirements for reserve, energy and ramping, the capability of each generator to provide each of the three services, and the overall least cost solution, the optimization software would also produce a breakdown of what proportion of the capacity of each unit was used to provide each of the three services in each trading period. Payment would then cover energy delivered x energy price + reserves delivered x reserve price + ramping delivered x ramping price. To the extent a unit would have made more money under some alternative optimal “unconstrained” dispatch (unconstrained, for example, with respect to transmission constraints), CMSC payments would be made.

E. 1X Ramp Rate, Multi-Interval Optimization, Ramping as Ancillary Service

Finally, the previous option could be implemented using the actual ramp rates and the multi-interval optimization algorithm.

III. Criteria

All six of the energy/ramping pricing schemes above will “work” in the sense that they will generate prices which generators will be paid. Choosing between them is a question of what criteria we want a pricing scheme to meet. The IESO proposes the following criteria. These criteria are not all simultaneously achievable, nor are they mutually exclusive. Trade-offs exist. The weighting factors that should be used to apportion the relative importance of these criteria

should be consistent with the overall goals of the Ontario market, and should generally be consistent with those used in market rule decisions that have already been made.

A. Efficiency

Prices serve as signals. It is natural that we should want those signals to tell us something about the costs and benefits of behaviour. In particular, the prices derived should, all other things equal:

1. produce a given quantity of electricity at minimum cost
2. allow consumers to curtail uses of electricity whose value to them falls short of the cost of providing that increment of power
3. incentivize the entry and exit of producers to enter the market with that mix of capacity and characteristics which is most valued by consumers in both the short run and the long run,
4. Incentivize changes in plant characteristics of existing plants whose social benefits exceed social costs

B. Fairness

Fairness is often easier to identify than to define. Broadly construed, fairness is an attribute that identical parties (in some respect) should be treated identically. Thus, all generators providing energy should receive the energy price. No generator who is needed should be compelled to provide energy at a loss. Note that even these two unobjectionable principles are partly in conflict. In any case, we will discuss for each of the options potential sources of unfairness whose ultimate impact will be left for the reader.

C. Reliability

The rules must allow the system to run reliably. Much care has been taken to ensure that a generator's willingness to follow dispatch instructions should be immediate and heedless of the financial consequences by ensuring that generators are essentially made whole so long as they follow those instructions.

Furthermore, an emphasis on system reliability should minimize the frequency with which operators need to take emergency action. Whenever emergency action is taken, there is the possibility of catastrophic failure.

Finally, the reliability criterion has a long-term component. System reliability should be a consequence of market actions, not out-of-market actions in the long run. If ramping resources are needed and are more expensive than non-ramping facilities, compensation schemes ought to provide sufficient return to incentivize the ramping facilities required.

D. Transparency

To the extent possible, price determination should be transparent. Transparency aids in the creation of contracts which in turn facilitate hedging. While the computer algorithms which produce prices can never be completely transparent, in general price should respond in predictable patterns as system conditions change. Identical conditions ought to produce identical prices. All derived prices should be explainable and sensible after the fact.

E. Robustness

From time to time, changes are required in the Ontario market. Temporary conditions ought not to justify a methodology which will be more difficult to change later. Thus, if we choose to ignore ramping costs and thereby make it more difficult to accommodate real ramping problems in the future then we will have been ill-served if ramping could be addressed easily today.

F. Cost

Without embarking on extensive cost-benefit analysis of every proposed change to the system, there may well be pricing methodologies requiring such extensive changes to the dispatch and settlement systems currently employed as to seriously call into question whether the Ontario market can support the cost.

IV. Evaluation

We are now prepared to analyze the proposed solutions against the criteria.

A. Status Quo: 12x Ramp, Myopic Optimization, No Minimum Load Constraints

1. Efficiency

If employed for actual system dispatch, 12X ramping and myopic optimization would require constant action on the part of system operators to maintain efficient and reliable dispatch. This should serve as a clear signal that the prices which derive from this algorithm do not support efficient (or for that matter feasible) dispatch.

The inefficiency of ignoring the ramp constraints takes several forms:

First, all ramping is managed through CMSC payments. CMSC payments create constrained-off payments which raise uplift prices inefficiently. Constrained-on units may not have appropriately modified their bids to reflect their proper status in the true merit order and therefore create an inefficient dispatch.

Second, to the extent that ramping requirements are actually binding and increase the overall price of energy (although perhaps decreasing it in some hours) failure to take this into account

makes the Ontario price too low relative to neighbouring jurisdictions. This will cause Ontario generators to export power rather than use it in Ontario.

Third, It should be noted as well that it may be difficult for ramping units to modify their bids in order to be compensated for ramping. First, it is difficult to do so in the absence of knowledge of the bid of others. Second, there may be many reasons to change bids – to modify commitment periods in order to recover start-up and no-load costs or to account for other nonconvexities, but there is only one instrument available to make that signal: the energy price⁸.

The effects of the status quo mechanism on energy prices is not clear. In some circumstances, this methodology clearly serves to understate the energy price. The obvious example of this phenomenon is when a slow-ramping unit has a low energy bid. The algorithm will presume that incremental energy could be pulled from this unit when, in fact, this unit is already ramping as fast as it can and cannot provide any more energy. In this case, the energy will have to come from a more expensive unit.

However, there are other times in which this algorithm clearly overstates the energy price. The clearest example of this is the period just before the ramp. In general, this period should have very low energy costs, since incremental energy consumed in this period reduces the level of the ramp. Through low prices we encourage energy usage sooner, reducing the necessity to ramp.

However, energy prices either above or below the true marginal costs of generators send improper signals to consumers. Baseload units are particularly affected. Their output is largely unaffected by ramping constraints – under normal conditions they will have been fully-loaded and thus incapable of ramping up in the morning whatever their technical capabilities. But if they are systematically under or over-compensated on weekday mornings through inappropriate signals it can send signals that either more or less baseload capacity is needed when such a conclusion is in fact unwarranted.

2. Fairness

Payment through CMSC payments certainly pays different amounts for ramping to units who have provided identical services to Ontario consumers. In this sense it is demonstrably unfair. In addition, there are probably inflexible units which receive CMSC constrained-off payments. These units are in no respect different than other inflexible units except that they have *higher* energy bids⁹. The units with the highest bids will receive the smallest payments for being constrained off. Had ramping been included in the optimization dispatch, these units would have received no payment at all.

⁸ Actually, of course, there are two signals: energy and reserve. While this may simplify the second problem of an inadequate number of signals, it undoubtedly complicates the first.

⁹ Actually, which units optimally receive the CMSC payments is quite complicated. If there is to be a unit which must be constrained off anyway for transmission constraints *and* which is inflexible, the least-cost solution is probably to get a double-barreled effect for only one CMSC payment. It is for this reason that precisely quantifying the value of ramping is liable to be quite difficult.

3. Reliability

The reliability of the *status quo* in the short run simply reflects the ability of the system operators to understand the morning ramp condition and meet it. As we understand it, there is currently no shortage of ways to meet the morning ramp in Ontario.

In the longer term, however, it is unclear whether or not there will be adequate ramping facilities in new units. In theory, a knowledge that if ramping becomes short a unit can move its bid to exercise “ramping market power” ought to bring forth ramping supply in new units. In practice, this theory suffers from the same issues as reliability in capacity markets: it is a highly volatile, “feast-or-famine” payment stream which may only work with long boom-and-bust cycles.

4. Transparency

The 12X ramping payment functions as the direct output of a computer algorithm. While the algorithmic results may not always be transparent, the lack of IESO intervention to produce pricing carries with it a sort of transparency. Furthermore, to the extent that similar conditions call forth similar prices, the system can be relied upon to produce stable prices so long as conditions are stable.

5. Robustness

The *status quo* works at the present time. If there was an actual shortage of ramping capacity, or if the units which could provide ramping capacity became so few that market power in ramping is a problem, then this particular method of compensations could probably not survive.

6. Cost

The *status quo* is clearly the least expensive solution to implement, as it has no incremental cost.

B. 1X Ramp Rate, Myopic Optimization, Minimum Load Constraint

1. Efficiency

Inclusion of actual ramping costs in the pricing algorithm has dramatic impacts on energy price, holding all offers constant. In order to provide ramping in the optimization program, large numbers of very expensive units are rushed into service. In essence, it is simply not possible to run the system according to myopic dispatch if every unit offers power at marginal costs.

But is energy really more valuable? The clear answer is that it is not. Evidence for this is that the actual dispatch achieves the required power at dramatically lower cost than the myopic dispatch. The myopic dispatch treats the system as if it is always in crisis during the morning ramp, when in fact the system operators have already taken actions to ensure that the system is not in crisis. Thus, the resulting signal is that energy is very valuable from six to ten a.m. even though, considering the full range of measures which could be taken to provide ramp, they can be accomplished (as the *status quo* solution shows) at very little cost.

Of course, these high myopic prices do not represent an economic equilibrium., In particular, high ramping prices in the morning would be expected to fill up intertie lines with imports, depressing the myopic price and cause generators capable of ramping to self-dispatch (by offering low bids in the early morning) to be available to provide ramp. This will depress the myopic prices as well.

At best, this system will replicate the multi-interval optimized dispatch. At worst, however, it will simply result in inefficient dispatch as inevitable errors creep into a decentralized system. This was the original vision of the Ontario market: wherever oddities in the myopic optimization cause regular and predictable arbitrage opportunities, the market participants are expected to modify their energy and reserve offers to take advantage of those opportunities. Since this has not been tested however, we have no evidence that this effect will work in practice.

It has been suggested that myopic pricing represents the “true” shadow price of energy. The fact that, even at best, generation arbitrage requires a skewing of bids to return the system to a competitive equilibrium indicates that this proposition cannot be true. Economic equilibrium is not myopic, but looks forward to cover all relevant foreseeable events.

The risk of myopic dispatch is the failure of generation action to return the system to the same dispatch as the multi-interval solution. If it does not, or if competition between generators is insufficient to constrain generators from reaping monopoly rents from ramping capacity, the resulting situation sends much too high a signal to consumers during the morning ramp. This will in turn cause consumers at the margin to refrain from consuming energy whose benefits exceed costs, *i.e.* causing allocative inefficiency. Further, it skews energy revenues in favour of those units which are already running before the morning ramp, *i.e.* baseload units. The more pronounced the effect, the more this methodology will favour incremental construction of baseload over mid-merit and peaking units. What it *will* do in this instance, however, is favour the construction of rapid-ramping mid-merit units to compete for the morning ramp, as well as quick-start units capable of ramping very quickly.

Compared with the *status quo*, this system will probably make fewer constrained-on payments, and therefore may be marginally more efficient from that perspective.

2. Fairness

This system, as noted above, probably favors baseload units over mid-merit units. In addition, it pays inflexible generators high energy prices even though it is their inflexibility which has caused the energy price to rise in the first place.

3. Reliability

Sending high prices for three hours a day to units which can get on and off relatively quickly undoubtedly helps reliability in the long run as potential entrants will no doubt design their units to take account of this effect. Of course, their doing so will take away the effect.

In the short run, the higher prices will undoubtedly cause more units to modify their bids to take advantage of the now decidedly less smooth price path. This in turn may well make the system more difficult to run in two respects. First, units modifying their bids rapidly cause the system optimizers to send out many more dispatch requests. Since each of these carries with it the possibility of error, system stability may suffer.

Second, disaggregation of dispatch decisions may themselves make the system more difficult to run in a stable fashion. Good utility practice has evolved in a regime in which changes to unit characteristics are relatively infrequent. These standard practices may prove inadequate when the underlying data vary widely to try and capture arbitrage opportunities.

Myopic dispatch pricing sends strong signals to construct units capable of ramping. Indeed, these signals are stronger than the optimal signals, just as the 12X signals are too weak. This suggests that this pricing mode may lead to more ramping capability than either a 12X or MIO pricing.

4. Transparency

The myopic dispatch with ramping constraints will probably result in a lower level of CMSC payments and for this reason is probably marginally more transparent than the *status quo*. However, after generators have full opportunities to arbitrage these prices, the resultant dispatch may be anything but transparent.

5. Robustness

This methodology, to work sensibly, requires several features. First, there should be no market power in the control of dispatch facilities. There must not be any units which can offer energy at very high prices knowing that they will be needed to meet the morning ramp. Second, the system needs to converge on a stable competitive solution. The conditions under which this will happen are unknown.

6. Cost

As we understand it, this system has little incremental cost to implement as it simply requires changing the ramp rate multiplier from 12 to 1 in the existing pricing software. The same could be said for minor modifications to this regime, e.g. 6X ramping. Implementation of minimum load constraints might require some incremental work, however.

C. 1X Ramp Rate, Multi-Interval Optimization, Minimum Load Constraint

1. Efficiency

By switching to a multi-interval optimization and including the actual ramp rates and minimum load constraints, the optimization algorithm should give an accurate indication of what energy is

really worth, for the simple reason that this algorithm is closest to the algorithm that system operators actually use. Consequently, the starting conditions in any interval ought to be reasonably close to the actual conditions that prevail (excepting transmission constraints, of course.)

This system will still make some constrained-on and –off payments to ramping and inflexible units, since it still has only one energy price to use to fulfill multiple objectives, but these constraint payments should be lower when the energy prices necessary to bring various units on and off are implemented. It should be noted, however, that this system may still see price spikes in the mornings – they will simply occur somewhat earlier and be spread over more hours. It is unclear whether the net energy prices will be higher or lower than under the preceding methodologies.

Most importantly, the aggregate price signal under this system is probably closest to being correct, sending the correct signals about the value of energy and therefore not skewing power towards any type of unit or in favour of imports or exports.

2. Fairness

This system should fairly compensate all generators, except to the effect that constraint payments still pay different generators different amounts to provide the same service.

3. Reliability

The fact that pricing matches dispatch ought to marginally increase the reliability of the system in two respects. First, the lower level of constraint payments ought to make generators more willing to follow instructions, though this effect should be quite minor. More importantly, this system removes incentives to skew bids to “game” the dispatch pricing algorithm. Thus, one would expect more stable bidding behavior, and, consequently, a somewhat more reliable system.

In the long run, while this pricing method will send signals to add ramping capacity to the system, it will only do so in a “boom-and-bust” fashion. As with all capacity expansion signals, it is an open question whether or not highly lumpy incentives will actually bring forth the right amount of capacity.

4. Transparency

This system probably has a high degree of transparency since the system operators will in general require few overrides of the dispatch for ramping purposes.

5. Robustness

This payment scheme can certainly survive transitions to other market arrangements, including day-ahead markets and especially location-based prices. Indeed, should this be implemented in

parallel with transmission constrained pricing, the pricing algorithm would precisely mirror the dispatch, sending quite accurate signals.

6. Cost

As we understand it, this system has already been implemented in dispatch. The precise alignment of top-up payments to generators dispatched out-of-merit for ramp with the constrained-on and –off payments due to transmission congestion will undoubtedly require further work.

D. 12X Ramp Rate, Myopic Optimization, No Minimum Load Constraint, Payment to Ramping Units

1. Efficiency

Two proposals have emerged to keep the present system of payment but to make additional payments to units which ramp. One proposal pays the difference between the energy prices from the 12x and 1x myopic dispatches, the other a fixed payment.

If the purpose of these payments is to compensate those who are constrained-on, the argument above shows that there is no particular justification for doing so. Constrained-on units choose their level of compensation, which need not be marginal cost. To the extent that the ramping problem is a recurrent, predictable matter, bids should converge to a level in which all rampers are fairly compensated.

Furthermore, neither proposal actually compensates generators for the value of ramping. The fixed payment is obviously arbitrary. The payment based on the difference between energy prices under two dispatch algorithms bears no necessary relationship to the value of ramping, either; it simply signals the difference in the energy price which can be attributed to the existence of ramping constraints. This is not the same thing as the value of ramping.

It is unclear whether the payments to be made are meant to replace CMSC payments. If a unit is constrained-on and is called on to ramp, the logic of CMSC payments would suggest that these payments should be reduced by the ramping payments made, if any. The problem with doing so is that it is unclear why a unit is being constrained on. Some are constrained on for ramp; others are constrained on to relieve transmission constraints. For many units it will be some combination of both.

Further, while these payments may raise the payments to ramping units, they will do nothing to reduce the constrained-off payments to inflexible units. Thus, while providing (perhaps) additional incentive to ramp, they also provide (inefficient) payments to those whose inflexibilities have caused a ramp problem.

Finally, changing one's level of generation is not always a response to a ramping constraint. Thus, to pay all changes in generation levels a fixed charge will compensate units for actions which are in fact unrelated to ramping constraints. When prices fall, some units will be asked to

turn down or turn off. It makes little sense to make extra payments to these units when it is in fact their own bids which caused them to be shut off. Thus, if any proposal like this is to be adopted, it should condition the payment on some notion, even an imprecise one, of the value of ramping in a particular interval.

2. Fairness

These proposals are fairer than the schemes which rely on CMSC payments since all units receive the same price for ramping. However, for this to be the case, ramping payments must be higher than CMSC payments – otherwise, different units will once again be paid different amounts to provide the same service.

Suppose two units have the same bid for energy and the same ramp rate. One is ordered to ramp down while the other is kept at minimum load. The ramping unit is paid under this system more than the one kept online at minimum load.

3. Reliability

In the short-run, this methodology should make little difference to the reliability of the system, since it is really just a change to the compensation methodology, not to system operations. In the longer run, providing extra payments to ramping units will probably give incentives for more ramping, but it is unclear that those expenses will be efficient, since there is no particular reason to believe either payment signal is efficient.

4. Transparency

A fixed payment for ramping is obviously transparent. The payment made on the difference in price between two alternative optimizations is obviously highly nontransparent. More problematic is the question of who is asked to ramp and why. Dispatch instructions will be unchanged, presumably.

5. Robustness

The fixed fee will undoubtedly require changes as the market develops, particular if they are not calculated net of any CMSC payments. Neither of these systems is well suited to operation in the case of a shortage of ramping facilities, although the variable payment scheme will at least give a rising incentive in that case.

6. Cost

Direct changes to the compensation scheme require changes to the settlement software. In addition, the difference compensation plan requires more resources to calculate a payment for ramping units.

E. 12X Ramp Rate, Myopic Optimization, No Minimum Load Constraint, Ramping as Ancillary Service

1. Efficiency

If ramping is a significant constraint on the system, the most efficient way to pay for it is to create an ancillary service for ramping and pay those who offer to solve the problem at least cost, taking into account their ability to fulfill other constraints, *e.g.*, energy and reserves. For an ancillary service scheme to work:

- The service provided must be competitive, *i.e.* there should not be market power in the provision of ramping.
- The service must be co-optimized, *i.e.* the provision of ramping must compete with provision of energy and reserves. Any choice to dispatch a unit in one way rather than another changes what can be done in all three dimensions. The optimization software needs to take that into account.
- The service must be optimized under a system in which the constraint is binding. Otherwise the price will always be zero.

The advantage of such a system is that it sends a precise signal in all dimensions. Energy, reserves, and ramp are all compensated according to their value to the system at any given time. The co-optimization ensures that few constrained-on or –off payments are made, although the fact that the system is still unconstrained for transmission constraints may mean that the units which would notionally provide, say, ramp, cannot do so because they are located on the wrong side of a transmission constraint. Consequently, some other units may receive constrained-on and –off payments to provide ramp even though ramping has nominally been co-optimized.

Running this scheme with the myopic optimization, however, still creates a mismatch between the generation which is optimal and the generation which comes from the system operator's dispatch. This mismatch in turn will cause a much larger level of constraint payments and is thus less efficient than the MIO optimization.

2. Fairness

Offering ramp as an ancillary service treats all units fairly. Units which perform the same value of service to the system will receive the same compensation.

3. Reliability

This payment system is equivalent to other payment systems which use myopic dispatch. Unit arbitrage in bidding may make the system more difficult to operate stably. Dynamically, by signaling the true value of ramping, new construction can choose a ramping capability which appropriate values the ability to ramp.

4. Transparency

As the number of co-optimized services rise, the ability to diagnose the exact reason why prices are as they are falls markedly.

5. Robustness

If implemented, the true value of ramping is continuously signaled. When ramping has little value, few payments will be made. When ramping has great value many payments will be made. Should other ancillary services be added in the future, the experience gained from adding one here should be valuable.

6. Cost

Setting up ramp as an ancillary service is bound to be a costly solution. Extensive changes must be made not only to the settlement software, but to bidding platforms as well. Training and testing make this the most expensive alternative in implementation costs.

F. 1X Ramp Rate, Multi-Interval Optimization, Minimum Load Constraints, Ramping as Ancillary Service

1. Efficiency

With the usual caveat that the presence of transmission constraints could potentially undo the efficiency of any system, this system should give the most efficient signals for three reasons:

- First, with three degrees of freedom to express willingness to provide service, generators have a better ability to signal their availability to provide ramp without compromising their ability to provide energy or reserves.
- Second, this system should have the lowest level of CMSC constrained-off payments, at least as far as ramping is concerned.
- Third, co-optimization of energy, reserves and ramping capability give the best feasible dispatch.

2. Fairness

Each service is provided no more or less than its marginal social value in payment, which clearly enhances fairness. The lower values of constrained-on payments also treat equivalent units equally.

3. Reliability

By having the dispatch align closely with actual system planning, this system should be quite reliable. In the long run, to the extent that ramping becomes a valuable commodity, this system sends appropriate signals.

4. Transparency

Same as the previous option.

5. Robustness

This system should be somewhat more robust than the ancillary service with myopic dispatch owing to the congruence between system planning and payment algorithms.

6. Cost

Implementation costs should be the same as the previous option.