

Оптимальный период записи контрольной точки для длительных вычислительных задач

А.В. Киселёв

НИИ «Квант»

В статье уточняются известные результаты для задач, длительность выполнения которых соизмерима со средним временем наработки на отказ вычислительной системы. Оптимальная длительность периода записи контрольной точки определяется на основе предложенных критериев максимизации показателей временной и энергетической эффективности. Исследованы зависимости показателей эффективности от параметра режима обработки — коэффициента перекрытия временных интервалов вычислений и записи контрольной точки.

Ключевые слова: оптимальная длительность периода записи контрольной точки, длительные вычислительные задачи, максимизация показателя временной эффективности, максимизация показателя энергетической эффективности.

1. Необходимость уточнения известных результатов для длительных вычислительных задач

Для крупномасштабных вычислительных задач, длительно использующих большое количество вычислительных узлов, обеспечение отказоустойчивости вычислений приобретает важное значение, поскольку время решения задачи может быть сравнимо и даже превышать время наработки на отказ, составляющее для современных суперкомпьютеров всего несколько часов [1-3]. Одним из основных способов обеспечения отказоустойчивости на пользовательском уровне является периодическое сохранение состояния вычислений (запись контрольной точки) для возможности последующего его восстановления в случае отказа. Запись контрольной точки (КТ), как и восстановление состояния вычислений после отказа, приводит к дополнительным временным и энергетическим затратам, при этом, очевидно, что чем чаще записывается КТ тем меньше затрат потребуются на повторное выполнение несохраненной вычислительной работы.

Определению оптимального периода записи контрольной точки по критериям минимизации времени решения задачи и (или) минимизации энергопотребления при ее решении посвящены работы [4-8]. В указанных работах приняты допущения, справедливые только в случае небольшой длительности вычислений ($\mathcal{T}_{\text{final}}$): вероятность отказа принимается равной $\frac{\mathcal{T}_{\text{final}}}{\mu}$, где μ – средняя наработка на отказ. Такое допущение справедливо при $\mathcal{T}_{\text{final}} \ll \mu$ и неприменимо при соизмеримых значениях $\mathcal{T}_{\text{final}}$ и μ . Для ряда вычислительных задач $\mathcal{T}_{\text{final}}$ не может быть оценена в принципе, например для случая обработки потока данных. Поэтому требуют уточнения приведенные в указанных работах модели, содержащие $\mathcal{T}_{\text{final}}$.

В настоящей статье вычисления рассматриваются как бесконечный периодический процесс, а оптимальный период записи контрольной точки определяется в результате решения оптимизационной задачи, критерий которой (максимизация временной или энергетической эффективности) сформирован на основе зависимостей между длительностями элементов структуры цикла.

2. Предлагаемая модель процесса вычислений с записью КТ

При разработке модели процесса вычислений с записью КТ и оценке ее параметров примем следующие допущения:

1. В качестве значений параметров модели, являющихся случайными величинами, будем использовать их математические ожидания.

- Отказы будем считать внезапными. В этом случае поток отказов считается простейшим, с экспоненциальной функцией распределения интервала времени между отказами и пуассоновским распределением количества отказов на интервале. При заданном среднем времени безотказной работы системы (средняя наработка на отказ) μ интенсивность потока отказов равна $1/\mu$. Для простейшего потока отказов вероятность возникновения отказа на интервале времени t зависит от длительности интервала и интенсивности потока отказов и равна $P_t = 1 - e^{-t/\mu}$. Момент возникновения отказа на заданном интервале времени равномерно распределен по длительности интервала.
- Оценивая значения параметров модели будем применять известное соотношение: $\forall |x| \ll 1 (e^x \approx 1 + x)$.

Для рассмотрения режимов блокирующей (во время записи КТ обработка не выполняется) и неблокирующей (во время записи КТ продолжаются вычисления) записи КТ воспользуемся приемом, использованным в работе [7]. Введем параметр режима обработки данных ω : $0 \leq \omega \leq 1$, смысл которого - коэффициент перекрытия временных интервалов обработки данных и записи контрольной точки. При $\omega = 0$ перекрытие отсутствует, что соответствует режиму блокирующей записи КТ. При $\omega = 1$ контрольная точка полностью формируется и записывается во время вычислений, – неблокирующий режим.

Процесс обработки потока данных имеет циклический характер поэтому модель процесса полностью описывается временной структурой цикла (рис. 1). Для анализа зависимостей между элементами цикла определим 3 временных интервала возможного возникновения отказа.

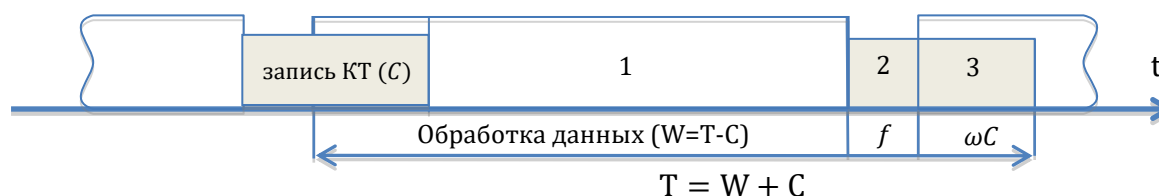


Рис. 1. Временная структура цикла обработки данных без учета отказов.

При отсутствии отказов цикл работы системы T включает временные интервалы обработки данных W и сохранения состояния системы C и равен периоду записи контрольной точки: $T = W + C$.

Длительность сохранения контрольной точки C включает длительность этапа формирования данных контрольной точки f и длительность этапа записи данных в файл $C - f$. Для обеспечения согласованного состояния сохраняемых данных на время f — формирования КТ — вычисления приостанавливаются. Для длительности обработки, выполняемой во время записи КТ, справедливо $\omega C \leq C - f$. Откуда следует ограничение $\omega \leq 1 - \frac{f}{C}$.

Если вся обработка выполняется во время записи КТ ($0 < W \leq C - f$), то максимальная длительность периода записи КТ при заданном ω составит: $T_{0 < \omega \leq 1, W = \omega C} = \omega C + C = C(1 + \omega)$. При $W < \omega C$ возникают простои в вычислениях из-за ожидания завершения сохранения состояния. Поэтому для практически значимых случаев соотношение между длительностью обработки и записи КТ составляет: $W \geq \omega C$, откуда следует ограничение для значений $T \geq C(1 + \omega)$. Нижнюю границу области определения T обозначим как $T_{inf} = C(1 + \omega)$.

В известных моделях [7] учет влияния отказов (\mathcal{T}) осуществляется в выражении для длительности решения задачи: $\mathcal{T}_{final} = \mathcal{T}_{final,ff} + \mathcal{T}$, где $\mathcal{T}_{final,ff}$ - длительность решения задачи без отказов. Поскольку в разрабатываемой модели вычисления рассматриваются как бесконечный периодический процесс и время решения задачи величина неопределенная, влияние отказов будем учитывать в выражении для длительности цикла параметром τ , равным вносимой отказами средней задержке в обработке данных. Задержка связана с восстановлением работоспособного состояния системы после отказа и восстановлением состояния вычислений – повторным выполнением вычислительной работы, результаты которой не были сохранены.

Обозначим длительность восстановления состояния — B и, где это необходимо, будем рассматривать длительность отдельных этапов восстановления — останов системы — D , чтение данных из набора данных контрольной точки и перезапуск системы — R : $B=D+R$.

Величина задержки зависит от момента возникновения отказа на длительности цикла: чем ближе момент отказа к моменту успешного завершения записи КТ, тем меньше длительность задержки.

Заметим, что отказ в момент восстановления состояния вычислений есть сложное событие XY , состоящее в одновременном наступлении следующих независимых (так как поток отказов без последствия) событий: X - отказ на некотором интервале с длительностью T_X цикла обработки; Y - отказ на длительности T_Y интервала задержки, связанной с восстановлением после данного отказа. Вероятность такого события $P(XY)=P(X)P(Y)$. Учитывая, что всегда $T_X \geq T_Y - B$ (за время T_Y выполняется восстановление состояния и повторное выполнение работ, выполненных до отказа на T_X), можно записать:

$$P(X) = \frac{T_X}{\mu}, P(Y) \leq \frac{B}{\mu} + \frac{T_X}{\mu}, P(XY) \leq \frac{T_X^2 + T_X B}{\mu^2}.$$

Поскольку $T_X \ll \mu$ и $B < \mu$, вероятность события $P(XY)$ пренебрежимо мала. И отказ во время восстановления после отказа будем считать невозможным событием.

С учетом вносимой отказом задержки длительность цикла равна: $T_o = T + \tau = W + C + \tau$.

Потребление энергии для разных интервалов цикла зависит от длительности выполняемой работы (τ_i) и мощности, потребляемой в процессе выполнения данного вида работы (e_i). Введем обозначения для мощностей следующих видов работ: вычисления (e_w), формирование и запись КТ (e_c), восстановление состояния системы после отказа (e_r). Мощность, потребляемую системой во время отказа обозначим как e_d . Кроме того, как и в [7], будем считать, что на всей длительности цикла существует постоянная составляющая энергопотребления E_{Π} с мощностью e : $E_{\Pi} = T_o e = (\tau + W + C)e$. Данное энергопотребление связано с поддержанием функционирования системы.

В Таблице 1 для каждого из 3-х возможных временных интервалов возникновения отказа приведены выражения для вероятности отказа (P_i), длительности временной задержки (τ_i), а также энергозатраты на восстановление процесса обработки после отказа (E_{ij}).

Таблица 1. Значения параметров модели для различных временных интервалов возникновения отказа

| № интервала отказа, i | Длительность интервала | Вероятность отказа на интервале, P_i | Задержка при отказе на интервале, τ_i | Описание компонентов формулы | Элементы энергозатрат на выполнение работы, E_{ij} |
|-------------------------|------------------------|--|---|--|--|
| 1 | $W - \omega C$ | $P_1 = \frac{W - \omega C}{\mu}$ | $\tau_1 = B + \omega C + \frac{1}{2}(W - \omega C) = B + \frac{1}{2}(W + \omega C)$ | B – восстановление системы после отказа | $E_{11} = De_d + Re_r$ |
| | | | | ωC – повторение обработки, выполненной во время записи КТ | $E_{12} = \omega C e_w$ |
| | | | | $\frac{1}{2}(W - \omega C)$ – средняя длительность обработки на интервале №1 до отказа; повторение обработки до отказа | $E_{13} = \frac{1}{2}(W - \omega C)e_w$ |
| 2 | $f = C - \omega C$ | $P_2 = \frac{f}{\mu} = \frac{C - \omega C}{\mu}$ | $\tau_2 = B + W + \frac{f}{2}$ | B – восстановление системы после отказа | $E_{21} = De_d + Re_r$ |
| | | | | W – повторная обработка данных | $E_{22} = We_w$ |

| | | | | | |
|---|------------|------------------------------|--|--|---|
| | | | $= B + W + \frac{C(1-\omega)}{2}$ | $\frac{f}{2} = \frac{C(1-\omega)}{2}$ – средняя длительность формирования данных КТ до отказа | $E_{23} = \frac{C(1-\omega)}{2} e_c$ |
| 3 | ωC | $P_3 = \frac{\omega C}{\mu}$ | $\tau_3 = B + W + f + \frac{1}{2}\omega C$ | B – восстановление системы после отказа | $E_{31} = De_d + Re_r$ |
| | | | $= B + W + C - \frac{1}{2}\omega C$ | W – обработка данных | $E_{32} = We_w$ |
| | | | | $f = C(1-\omega)$ – формирование КТ | $E_{33} = C(1-\omega)e_c$ |
| | | | | $\frac{1}{2}\omega C$ – средняя длительность обработки до отказа на интервале № 3; повторение обработки до отказа, при записи КТ | $E_{34} = \frac{1}{2}\omega C(e_w + e_c)$ |

Среднюю длительность задержки τ в результате отказа будем оценивать как математическое ожидание длительности задержки при возникновении отказа на разных интервалах цикла обработки: $\tau = P_1\tau_1 + P_2\tau_2 + P_3\tau_3$

С учетом введенных обозначений длительность цикла определяется формулой:

$$T_o = W + C + P_1\tau_1 + P_2\tau_2 + P_3\tau_3 \quad (1)$$

Подставив в формулу (1) значения из Таблицы 1, и, выполнив преобразования, получим выражение для длительности цикла обработки потока данных с учетом отказов:

$$T_o(W, \omega) = W + C + \frac{W-\omega C}{\mu} \left(B + \frac{1}{2}(W + \omega C) \right) + \frac{C-\omega C}{\mu} \left(B + W + \frac{C(1-\omega)}{2} \right) + \frac{\omega C}{\mu} \left(B + W + C - \frac{1}{2}\omega C \right) = W + C + \frac{1}{\mu} \left(B(W + C) + \frac{1}{2}(W + C)^2 - \omega C \left(B + \frac{\omega C}{2} \right) \right).$$

С учетом $W + C = T$, получаем:

$$T_o(T, \omega) = \frac{1}{2\mu} (T^2 + 2(B + \mu)T - \omega C(2B + \omega C)) \quad (2)$$

Из $\tau = T_o - T$ получим выражение для τ :

$$\tau = \frac{1}{2\mu} (T^2 + 2BT - \omega C(2B + \omega C))$$

Энергозатраты на каждый цикл обработки без учета отказов составляют:

$$E_n = We_w + Ce_c + \omega Ce_w + eT.$$

Средние энергозатраты на повторное выполнение работ в случае отказов — ε определяются как математическое ожидание энергозатрат на выполнение видов работ E_i , элементы энергозатрат которых приведены в Таблице 2:

$$\varepsilon = \sum_{i=1}^3 P_i E_i = \sum_{i=1}^3 \sum_{j=1}^k P_i E_{ij}$$

Подставив в данную формулу значения из Таблицы 1 и, выполнив равносильные преобразования, получим:

$$\varepsilon = \frac{W-\omega C}{\mu} \left(De_d + Re_r + \omega Ce_w + \frac{1}{2}(W - \omega C)e_w \right) + \frac{C-\omega C}{\mu} \left(De_d + Re_r + We_w + \frac{C(1-\omega)}{2} e_c \right) + \frac{\omega C}{\mu} \left(De_d + Re_r + We_w + C(1-\omega)e_c + \frac{1}{2}\omega C(e_w + e_c) \right) = \frac{1}{2\mu} (2(De_d + Re_r)(W + C) -$$

$$2\omega C(De_d + Re_r) + e_w(W + C)^2 - C^2(e_w - e_c) = \frac{1}{2\mu} (e_w T^2 + 2(De_d + Re_r)T - 2\omega C(De_d + Re_r) - C^2(e_w - e_c)).$$

Общий объем энергозатрат на цикле обработки с учетом отказов составляет:

$$E_o = E_n + \tau e + \varepsilon = W(e_w + e) + C(e_c + e) + \omega C e_w + \tau e + \varepsilon$$

Подставив в данную формулу значения для τ и ε , представим E_o как полином от T :

$$E_o(T, \omega) = W(e_w + e) + C(e_c + e) + \omega C e_w + \frac{e}{2\mu} (T^2 + 2(D + R)T - 2\omega C(D + R) - \omega^2 C^2) + \frac{1}{2\mu} (e_w T^2 + 2(De_d + Re_r)T - 2\omega C(De_d + Re_r) - C^2(e_w - e_c)) = eT + (T - C)e_w + C e_c + \omega C e_w + \frac{1}{2\mu} ((e_w + e)T^2 + 2(D(e_d + e) + R(e_r + e))T - 2\omega C(D(e_d + e) + R(e_r + e)) - C^2(e_w - e_c + \omega^2 e)) = \frac{1}{2\mu} (T^2(e + e_w) + 2T(D(e_d + e) + R(e_r + e)) + \mu(e_w + e)) - 2\omega C(D(e_d + e) + R(e_r + e)) - e_w C(C + 2\mu + \omega C)(1 - \omega) + e_c C(C + 2\mu) - \omega^2 C^2(e_w + e)).$$

Оптимальное значение периода записи КТ будем определять по критерию максимизации следующих показателей:

- временной эффективности $F_t(T, \omega) = \frac{W+\omega C}{T_o} = \frac{T-C(1-\omega)}{T_o}$ (характеризует эффективность использования ресурса – времени вычислений, для выполнения вычислительной работы на длительности цикла обработки);

- энергетической эффективности $F_e(T, \omega) = \frac{W+\omega C}{E_o} = \frac{T-C(1-\omega)}{E_o}$ (характеризует эффективность использования ресурса – энергии, для выполнения вычислительной работы на длительности цикла обработки).

3. Исследование зависимости показателя временной эффективности от параметра режима обработки данных

Рассмотрим зависимость показателя временной эффективности $F_t(T, \omega)$ от параметра ω :

$$F_t(T, \omega) = \frac{2\mu(T-C(1-\omega))}{T^2+2(B+\mu)T-\omega C(2B+\omega C)} = \frac{2\mu C\omega+2\mu T-2\mu C}{-C^2\omega^2-2BC\omega+T^2+2(B+\mu)T}.$$

Найдем производную $F_t(T, \omega)$ по ω :

$$\frac{dF_t(T, \omega)}{d\omega} = \frac{C^2\omega(4\mu T-4\mu C+2\mu C\omega)+2BC(2\mu T-2\mu C)+2\mu C(T^2+2(B+\mu)T)}{(\omega(-C^2\omega-2BC)+T^2+2(B+\mu)T)^2} = \frac{2C^2\omega(4\mu(T-C)+2\mu C\omega)+4BC\mu(T-C)+2\mu C(T^2+2(B+\mu)T)}{(T^2+2(B+\mu)T-\omega C(C\omega+2B))^2} = \frac{2C^2\omega(4\mu W+2\mu C\omega)+4BC\mu W+2\mu C(T^2+2(B+\mu)T)}{(2T\mu+(T-\omega C)(T+\omega C+2B))^2}$$

Поскольку $T - \omega C > 0$, числитель и знаменатель дроби положительны при $0 \leq \omega \leq 1$ и функция $F_t(T, \omega)$ возрастает по ω на всей области определения.

4. Исследование зависимости показателя энергетической эффективности от параметра режима обработки данных

Рассмотрим зависимость показателя энергетической эффективности $F_e(T, \omega)$ от ω .

$$F_e(T, \omega) = \frac{T-C(1-\omega)}{E_o} = 2\mu(T - C(1 - \omega)) / (T^2(e + e_w) + 2T(D(e_d + e) + R(e_r + e)) + \mu(e_w + e) - 2\omega C(D(e_d + e) + R(e_r + e)) - e_w C(C + 2\mu + \omega C)(1 - \omega) + e_c C(C + 2\mu) - \omega^2 C^2(e_w + e)) = (2\mu C\omega + 2\mu T - 2\mu C) / (-\omega^2 C^2 e - 2C(D(e_d + e) + R(e_r + e)) - \mu e_w)\omega + T^2(e + e_w) + 2T(D(e_d + e) + R(e_r + e)) + \mu(e_w + e) + C(e_c - e_w)(C + 2\mu))$$

Найдем производную функции $F_e(T, \omega)$ по ω :

$$\frac{dF_e(T, \omega)}{d\omega} = 2C\mu (C e_w(2T - 2C + C\omega) + 2(D(e_d + e) + R(e_r + e)) - \mu e_w)(T - C) + T^2(e + e_w) + 2T(D(e_d + e) + R(e_r + e)) + \mu(e_w + e) + C(e_c - e_w)(C + 2\mu) / (-\omega C(C e_w + 2(D(e_d + e) + R(e_r + e)) - \mu e_w))$$

$$R(e_r + e) - \mu e_w)) + T^2(e + e_w) + 2T(D(e_d + e) + R(e_r + e) + \mu(e_w + e)) + C(e_c - e_w)(C + 2\mu)^2 = 2C\mu \left(C e \omega (2W + C \omega) + 2W(D(e_d + e) + R(e_r + e)) + 2\mu e_w(T - W) + T^2 e + 2T(D(e_d + e) + R(e_r + e)) + C e_c(C + 2\mu) + 2\mu e(T - C) + e_w(T^2 - C^2) \right) / \left(e(T^2 - C^2 \omega^2) + 2(D(e_d + e) + R(e_r + e))(T - \omega C) + 2\omega C \mu e_w + e_w(T^2 - C^2) + 2\mu e_w(T - C) + 2T\mu e + C e_c(C + 2\mu) \right)^2$$

Поскольку $T > W, T > C, T - \omega C > 0$, числитель и знаменатель дроби положительны при $0 \leq \omega \leq 1$ и функция $F_e(T, \omega)$ возрастает по ω на всей области определения.

5. Оптимальная длительность периода записи КТ по критерию максимизации показателя временной эффективности

Исследование функции $F_t(T, \omega) = \frac{2\mu(T-C(1-\omega))}{T^2+2(B+\mu)T-\omega C(2B+\omega C)}$ показало, что при $0 \leq \omega \leq \frac{2\mu+2B+C}{2\mu+2(2B+C)}$ функция $F_t(T, \omega)$ достигает максимума в T_t , являющемся корнем уравнения: $\frac{dF_t(T, \omega)}{dT} = 0 \Leftrightarrow \frac{T(2C(1-\omega)-T)+2(B+\mu)C(1-\omega)-\omega C(2B+\omega C)}{(T^2+2(B+\mu)T-2\omega CB-\omega^2 C^2)^2} = 0 \Leftrightarrow T(2C(1-\omega) - T) + 2(B + \mu)C(1 - \omega) - \omega C(2B + \omega C) = 0$.

Так как $0 \leq \omega \leq \frac{2\mu+2B+C}{2\mu+2(2B+C)} \Rightarrow 2C(1-\omega)(\mu+B+C) - C(2\omega B + C) \geq 0$, то корень уравнения:

$$T_t = \sqrt{2C(1-\omega)(\mu+B+C) - C(2\omega B + C)} + C(1-\omega) \quad (3)$$

Таким образом, при $0 \leq \omega \leq \frac{2\mu+2B+C}{2\mu+2(2B+C)}$ функция $F_t(T, \omega)$ в точке T_t имеет максимум $F_t(T_t, \omega)$.

Для случая блокирующей записи контрольной точки ($\omega = 0$) получаем:

$$T_{t|\omega=0} = \sqrt{2C(\mu+B+C) - C^2 + C} \quad (4)$$

Из условия $T_t \geq T_{inf} \Leftrightarrow \sqrt{2C(1-\omega)(\mu+B+C) - C(2\omega B + C)} + C(1-\omega) \geq C(1+\omega)$ находим ограничение для ω : $\frac{\sqrt{4C(2\mu+2B+C)+(\mu+2B+C)^2} - (\mu+2B+C)}{4C} \geq \omega$. При выборе значения ω , следует также учитывать ограничение $0 \leq \omega \leq 1 - \frac{f}{C}$. Объединяя оба условия, получаем:

$$0 \leq \omega \leq \min \left\{ 1 - \frac{f}{C}, \frac{\sqrt{4C(2\mu+2B+C) + (\mu+2B+C)^2} - (\mu+2B+C)}{4C} \right\} \quad (5)$$

Так как $\frac{\sqrt{4C(2\mu+2B+C)+(\mu+2B+C)^2} - (\mu+2B+C)}{4C} \leq \frac{2\mu+2B+C}{2\mu+2(2B+C)}$ максимум $F_t(T_t, \omega)$ в точке T_t существует при любом ω , удовлетворяющем условию (5).

6. Оптимальная длительность периода записи КТ по критерию максимизации показателя энергетической эффективности

Найдем значение $T = T_e$, при котором достигается максимум $F_e(T, \omega)$:

$$T_e = \arg \max_{(T)} F_e(T, \omega).$$

Функция $F_e(T, \omega) = \frac{W+\omega C}{E_o} = \frac{T-C(1-\omega)}{E_o} = 2\mu(T-C(1-\omega)) / \left(T^2(e + e_w) + 2T(D(e_d + e) + R(e_r + e) + \mu(e_w + e)) - 2\omega C(D(e_d + e) + R(e_r + e)) - e_w C(C + 2\mu + \omega C)(1 - \omega) + e_c C(C + 2\mu) - \omega^2 C^2(e_w + e) \right)$ достигает максимума при T , являющемся решением уравнения $\frac{dF_e(T, \omega)}{dT} = 0$, где $\frac{dF_e(T, \omega)}{dT} = 2\mu \left((e + e_w)x(2C(1-\omega) - x) + 2(D(e_d + e) + R(e_r + e) + \mu(e_w + e))C(1-\omega) - 2\mu C(e_w - e_c) - 2\omega C((D + R)e + De_d + Re_r - \mu e_w) - C^2(\omega^2 e + e_w - e_c) \right) /$

$$\left(x((e + e_w)x + 2(D(e_d + e) + R(e_r + e) + \mu(e_w + e))) - 2\mu C(e_w - e_c) - 2\omega C((D + R)e + De_d + Re_r - \mu e_w) - C^2(\omega^2 e + e_w - e_c)\right)^2$$

Решая уравнение $\frac{dF_e(T, \omega)}{dT} = 0$ найдем значение T_e : $\frac{dF_t(T, \omega)}{dT} = 0 \Leftrightarrow (e + e_w)x(2C(1 - \omega) - x) + 2(D(e_d + e) + R(e_r + e) + \mu(e_w + e))C(1 - \omega) - 2\mu C(e_w - e_c) - 2\omega C((D + R)e + De_d + Re_r - \mu e_w) - C^2(\omega^2 e + e_w - e_c) = 0 \Rightarrow$

$$T_e = \sqrt{C \left(\frac{2(D(e_d + e) + R(e_r + e))(1 - 2\omega) + (2\mu + C)(e_c + e(1 - \omega)) - \omega C e_w(1 - \omega)}{(e + e_w)} - \omega C \right)} + C(1 - \omega) \quad (6)$$

В блокирующем режиме: $T_e = \sqrt{C \frac{2(D(e_d + e) + R(e_r + e)) + (2\mu + C)(e_c + e)}{(e + e_w)}} + C$.

Если полагать что энергозатраты на все виды работ одинаковые ($e_s = e_k = e_p$), данная формула сводится к формуле (4):

$$T_e|_{e_s=e_k=e_p} = \sqrt{2C(D + R) + C(2\mu + C)} + C = \sqrt{2C(B + \mu + C) - C^2} + C$$

7. Анализ полученных результатов

Для наглядности сравнения результатов, полученных в настоящей работе, с ранее известными, представим их в Таблице 2.

Таблица 2. Сравнение полученных результатов с ранее известными

| Работа, в которой приведен результат | Оптимальный период записи КТ T_t по критерию максимизации показателя временной эффективности $F_t(T, \omega)$ | Примечание |
|--------------------------------------|--|---|
| Новая модель | $\sqrt{2C(1 - \omega)(\mu + B + C) - C(2\omega B + C)} + C(1 - \omega)$ | Формула получена с учетом большой длительности вычислений |
| [5] (Young) | $\sqrt{2C\mu} + C$ | Не учитываются длительность восстановления после отказа и возможность неблокирующей записи КТ |
| [6] (Daly) | $\sqrt{2C(\mu + B)} + C$ | Не учитывается возможность неблокирующей записи КТ |
| [7] (Dongarra) | $\sqrt{2C(1 - \omega)(\mu - B - C\omega)}$ | Формула применима только при длительности решения задачи $T_{final} \ll \mu$ |
| | Оптимальный период записи КТ T_e по критерию максимизации показателя энергетической эффективности $F_e(T, \omega)$ | |

| | | |
|----------------|---|--|
| Новая модель | $\sqrt{C \left(\frac{2(D(e_d+e)+R(e_r+e))(1-2\omega)+(2\mu+C)(e_c+e(1-\omega))-\omega C e_w(1-\omega)}{(e+e_w)} - \omega C \right) + C(1-\omega)}$ | Формула получена с учетом большой длительности вычислений |
| [8] (El-Sayed) | $\sqrt{2C\mu \frac{e_c}{e_w}}$ | Не учитываются энергозатраты на восстановление после отказа и на запись КТ |
| [7] (Dongarra) | — | Зависимость в явном виде не получена |

При сравнении моделей использовались численные данные, приведенные в работе [7]. Исследовались следующие зависимости:

- оптимальной длительности периода записи КТ от средней наработки на отказ (рис. 2), от длительности записи КТ (рис. 3), от значения параметра режима обработки (рис. 4);
- показателей эффективности от значения параметра обработки (рис. 5) и длительности периода записи КТ (рис. 6);
- показателей эффективности длительности периода записи КТ при разных значениях параметра обработки (рис. 7);
- оптимальной длительности периода записи КТ от соотношения энергозатрат на вычисления и запись КТ (рис. 8);
- показателя энергоэффективности от соотношения энергозатрат на вычисления и запись КТ (рис. 9).

Анализ формул, приведенных в Таблице 2, и графиков зависимостей параметров моделей (рис. 2-4,6) показал что результаты, полученные для новой модели, согласуются с результатами, полученными в работах [4-8].

Из графиков, приведенных на рис. 6, следует, что значения оптимальной длительности периода записи КТ для разных критериев оптимизации (по времени, по энергии) существенно отличаются; вычисляемые по новым формулам значения длительности периода записи КТ (T_t, T_e), обеспечивают лучшие значения показателей эффективности среди всех рассмотренных моделей для критериев, введенных с учетом специфики рассматриваемого класса задач.

Из графиков на рис. 7 следует, что для достижения максимального значения показателя временной эффективности необходимо (с учетом ограничений (5)) выбирать наибольшее из допустимых значений параметра обработки и длительность периода обработки, вычисляемую по формуле (3).

7. Выводы

Приведенные в статье формулы для вычисления оптимальных значений параметра режима обработки и длительности периода записи контрольной точки. Критерием оптимальности является максимизация значений введенных показателей временной и энергетической эффективности. Полученные результаты дополняют и уточняют ранее известные для случая длительных вычислительных задач.

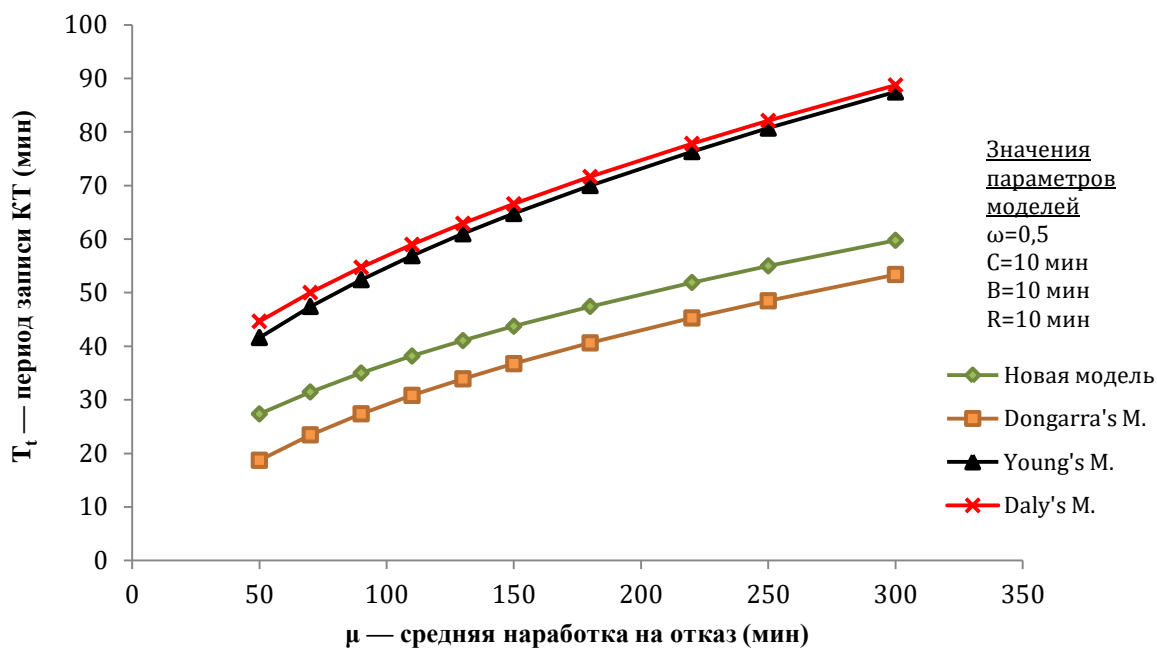


Рис. 2. Зависимость оптимальной длительности периода записи КТ от средней наработке на отказ

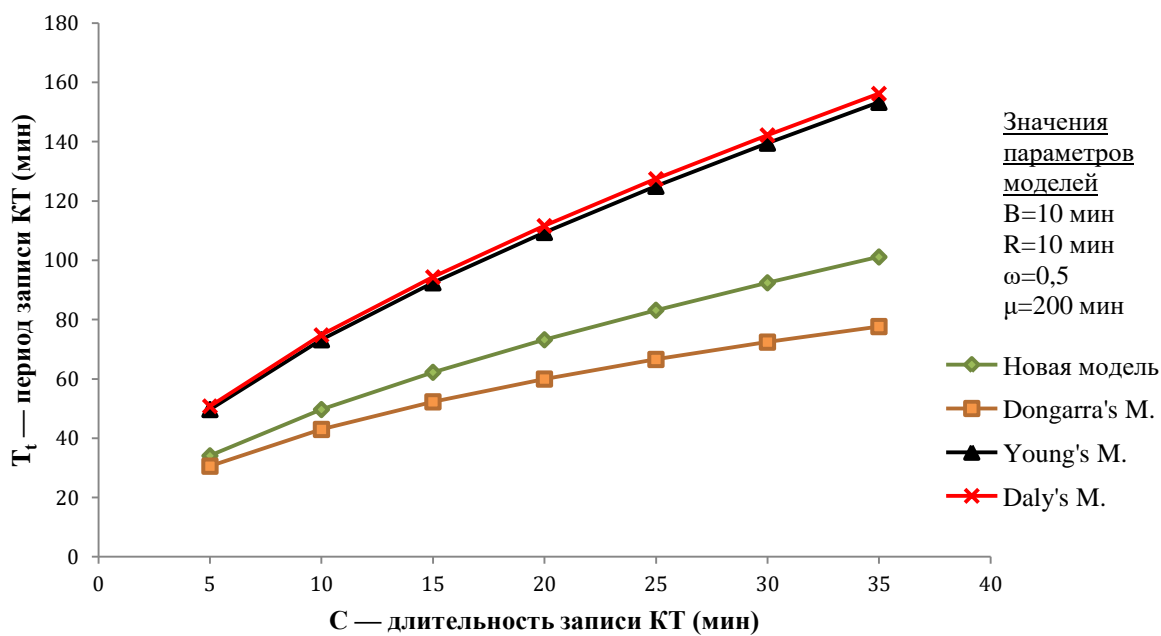


Рис. 3. Зависимость оптимальной длительности периода записи КТ от длительности записи КТ

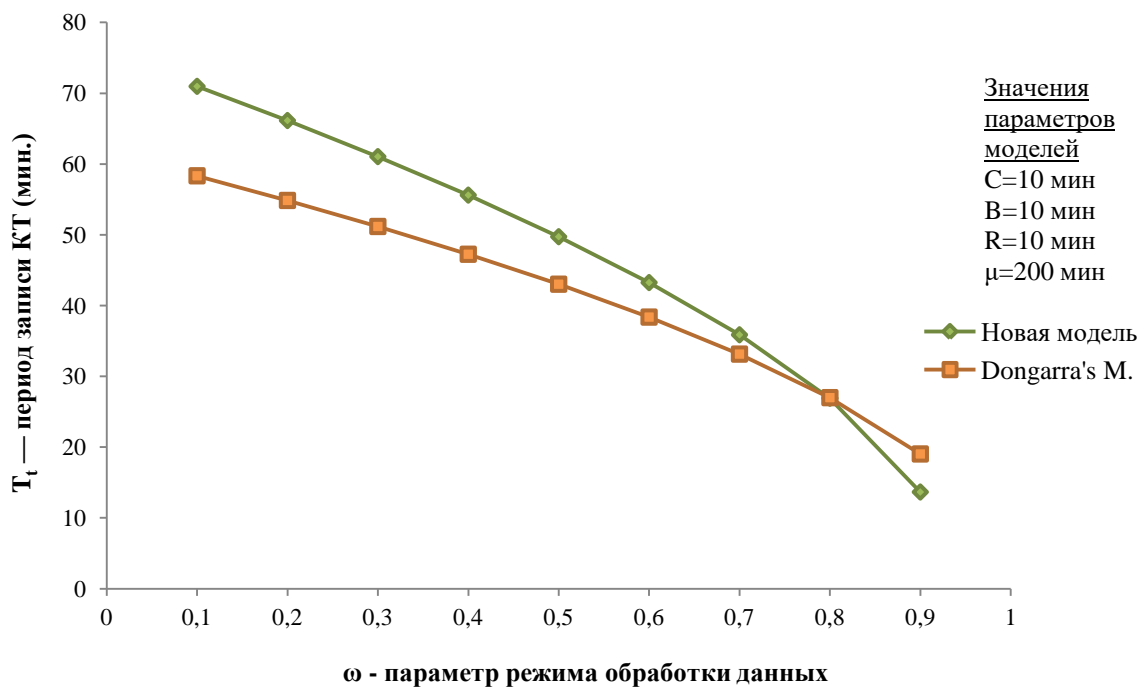


Рис. 4. Зависимость оптимальной длительности периода записи КТ от параметра режима обработки

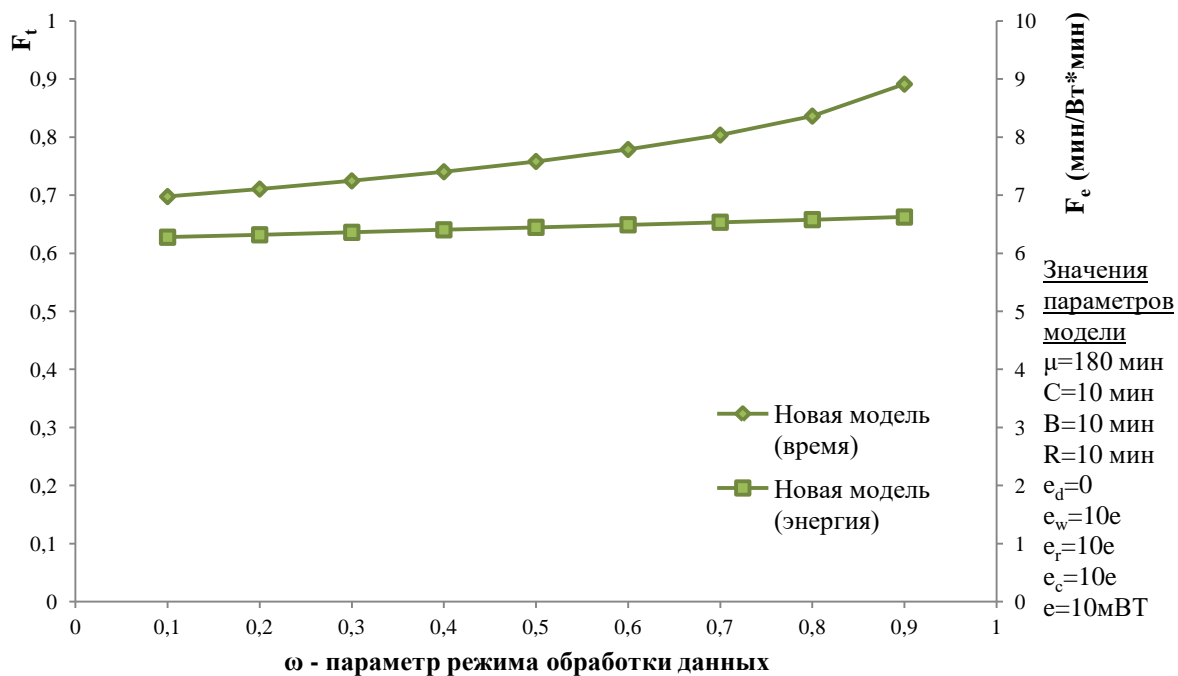


Рис. 5. Зависимость показателей эффективности от параметра режима обработки (ω)

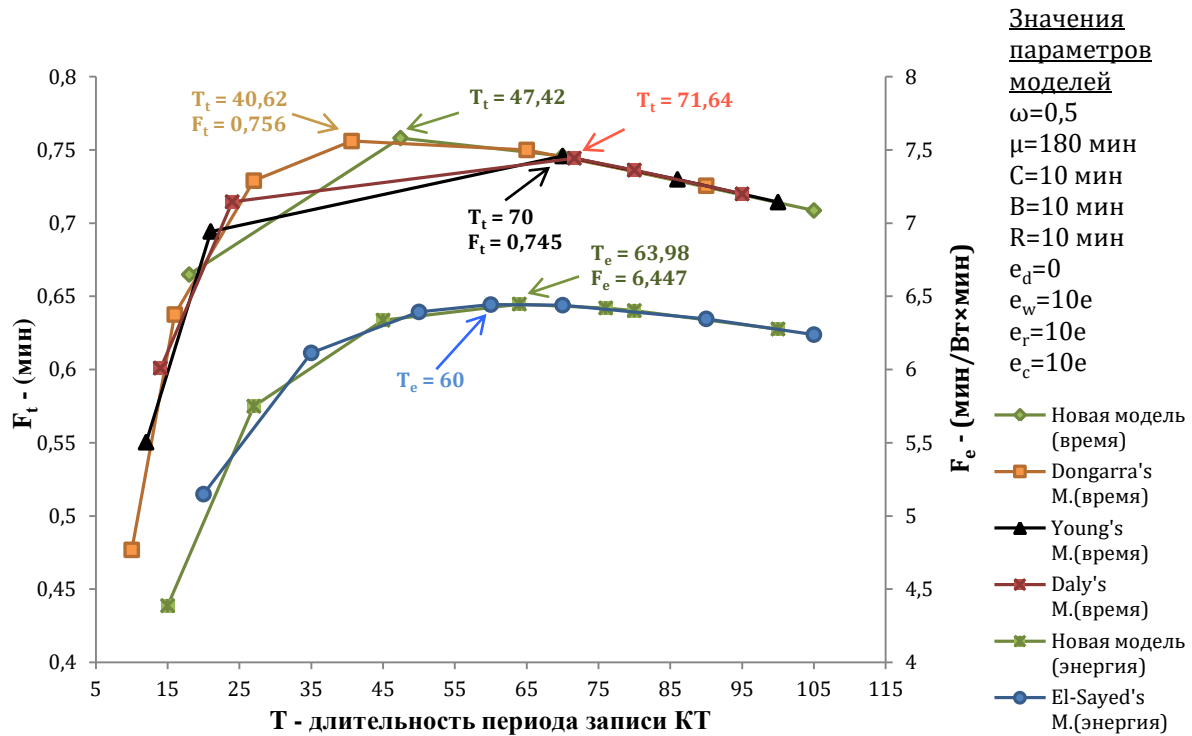


Рис. 6. Зависимости показателей эффективности от длительности периода записи КТ для разных моделей

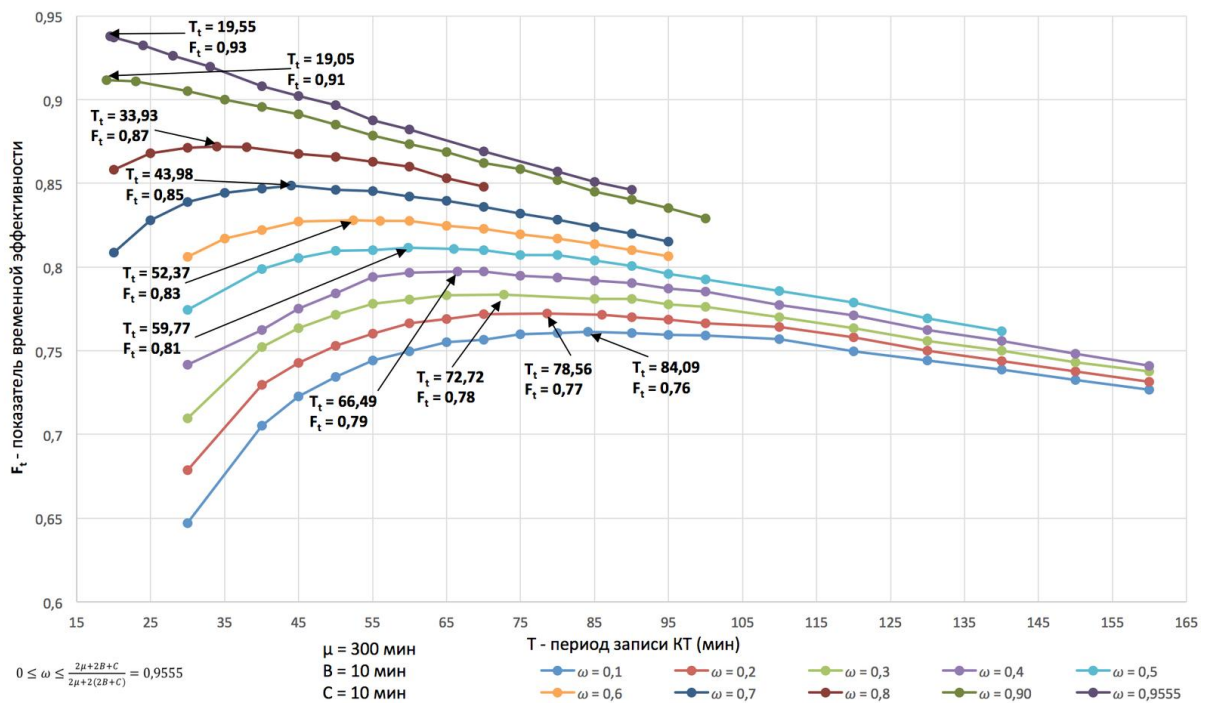


Рис. 7. Зависимость показателя временной эффективности от длительности периода записи КТ для разных значений ω

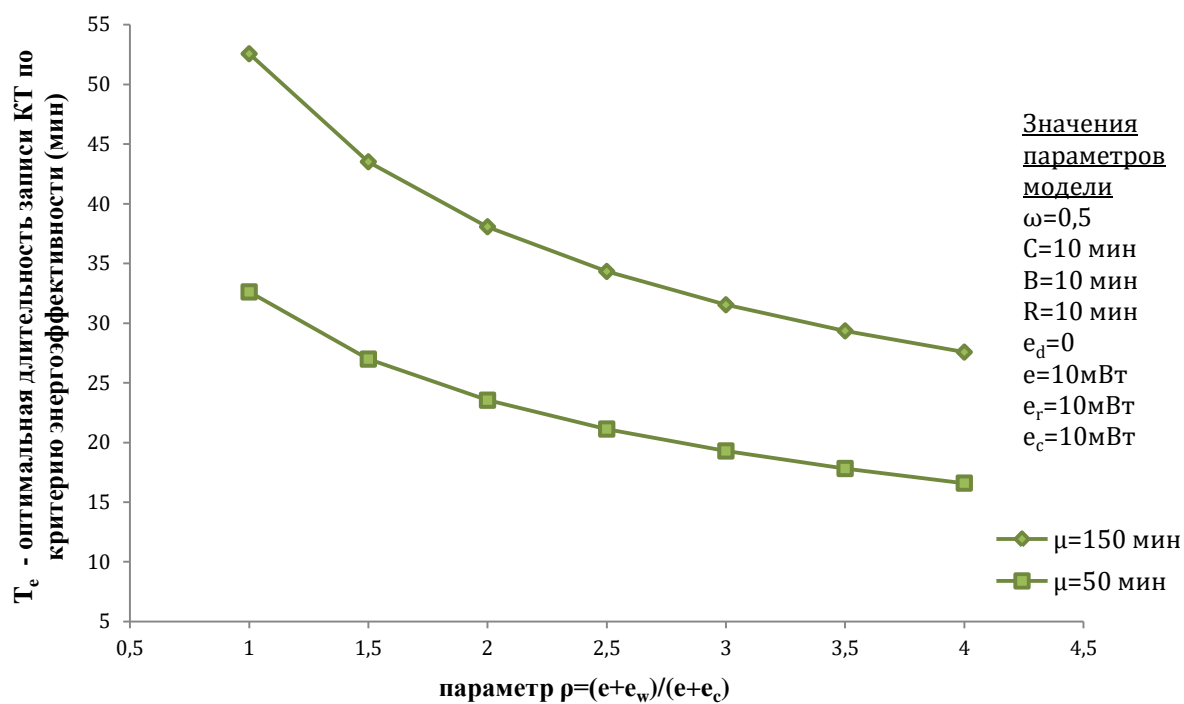


Рис. 8. Зависимость оптимальной длительности периода записи КТ (T_e) от соотношения энергозатрат на вычисления и запись КТ

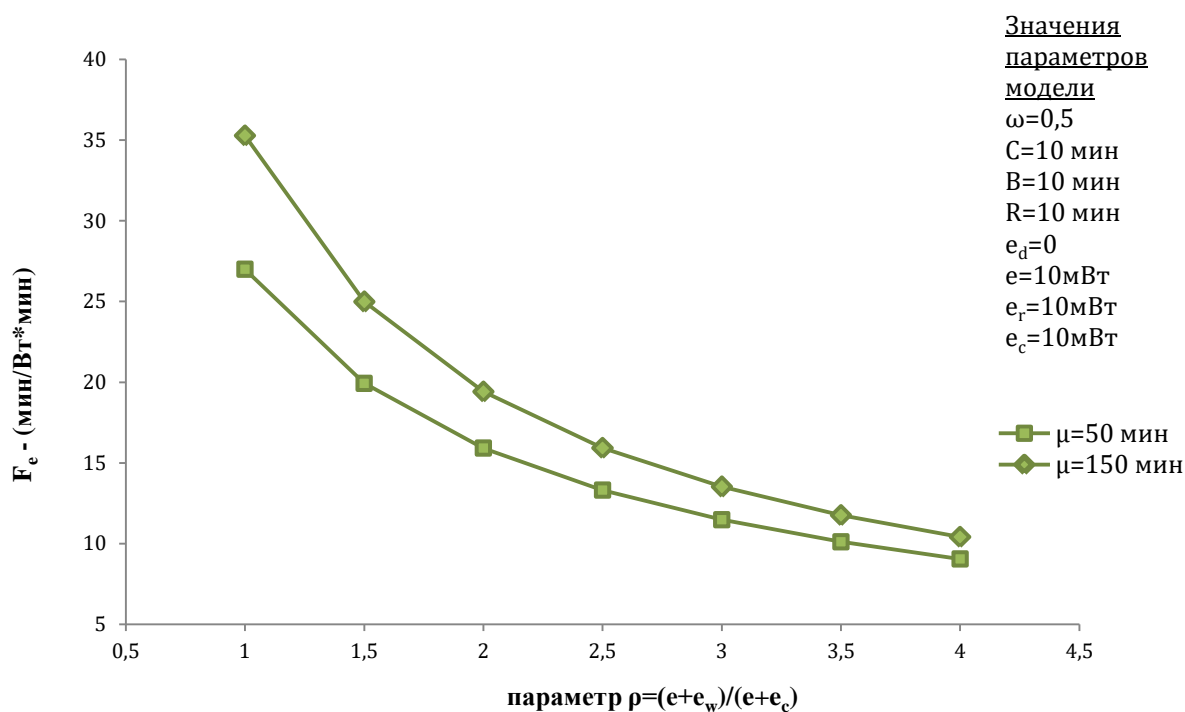


Рис. 9. Зависимость показателя энергоэффективности F_e от соотношения энергозатрат на вычисления и запись КТ

Литература

1. Бондаренко А., Якобовский М. Обеспечение отказоустойчивости высокопроизводительных вычислений с помощью локальных контрольных точек // Вестник Южно-Уральского государственного университета. Серия "Вычислительная математика и информатика". — 2014. — Т. 3, № 3. — С. 20–36.
2. D. Reed. High-end computing: The challenge of scale. Director's Colloquium, Los Alamos National Laboratory, May 2004.
3. Hsu, C.-H. A power-aware run-time system for high-performance computing / C.- H. Hsu, W.-C. Feng. // Proceedings of SC|05: The ACM/IEEE International Conference on High-Performance Computing, Networking, and Storage (Seattle, Washington USA November 12 – 18, 2005). — IEEE Press, 2005. — P. 1–9.
4. G. Zheng, X. Ni, and L. V. Kale', "A scalable double in-memory checkpoint and restart scheme towards exascale," in Dependable Systems and Networks Workshops (DSN-W), 2012.
5. J. W. Young, "A first order approximation to the optimum checkpoint interval," Comm. of the ACM, vol. 17, no. 9, pp. 530–531, 1974.
6. J. T. Daly, "A higher order estimate of the optimum checkpoint interval for restart dumps," FGCS, vol. 22, no. 3, pp. 303–312, 2004.
7. G. Aupy, A. Benoit, T. Herault, Y. Robert, and J. Dongarra. Optimal checkpointing period: Time vs. energy. CoRR, abs/1310.8456, 2013.
8. El-Sayed, N., Schroeder, B. To Checkpoint or Not to Checkpoint: Understanding Energy-Performance-I/O Tradeoffs in HPC Checkpointing / IEEE International Conference on Cluster Computing (CLUSTER 2014). Madrid, Spain, 22-26 September 2014. — IEEE Press, 2014. — P. 93 – 102.

Optimal checkpointing period for long duration computational tasks

Andrey Kiselev

Research and Development Institute "Kvant"

In the article refines known results for the case of tasks, the duration of the computation of which is commensurate with the Mean Time Between Failures of the computer system. Optimal checkpointing period is based on the proposed of maximization indexes of time-effectiveness and energy-effectiveness of computation criteria. Researched the dependence the indexes of time-effectiveness and energy-effectiveness from the parameter of processing mode - the coefficient of overlapping time intervals of the computing and checkpointing.

Keywords: optimal checkpointing period, long duration computational tasks, maximization indexes of time-effectiveness, maximization indexes of energy-effectiveness

References

1. A.A. Bondarenko, M.V. Iakobovski. Fault tolerance for by using local checkpoints // Bulletin of the South Ural State University Series "Computational Mathematics and Software Engineering" 2014, vol. 3, no. 3, pp. 20–36.
2. D. Reed. High-end computing: The challenge of scale. Director's Colloquium, Los Alamos National Laboratory, May 2004.
3. Hsu, C.-H. A power-aware run-time system for high-performance computing / C.- H. Hsu, W.-C. Feng. // Proceedings of SC05: The ACM/IEEE International Conference on High-Performance Computing, Networking, and Storage (Seattle, Washington USA November 12 – 18, 2005). — IEEE Press, 2005. — P. 1–9.
4. G. Zheng, X. Ni, and L. V. Kale', "A scalable double in-memory checkpoint and restart scheme towards exascale," in Dependable Systems and Networks Workshops (DSN-W), 2012.
5. J. W. Young, "A first order approximation to the optimum checkpoint interval," Comm. of the ACM, vol. 17, no. 9, pp. 530–531, 1974.
6. J. T. Daly, "A higher order estimate of the optimum checkpoint interval for restart dumps," FGCS, vol. 22, no. 3, pp. 303–312, 2004.
7. G. Aupy, A. Benoit, T. Herault, Y. Robert, and J. Dongarra. Optimal checkpointing period: Time vs. energy. CoRR, abs/1310.8456, 2013.
8. El-Sayed, N., Schroeder, B. To Checkpoint or Not to Checkpoint: Understanding Energy-Performance-I/O Tradeoffs in HPC Checkpointing / IEEE International Conference on Cluster Computing (CLUSTER 2014). Madrid, Spain, 22-26 September 2014. — IEEE Press, 2014. — P. 93 – 102.H